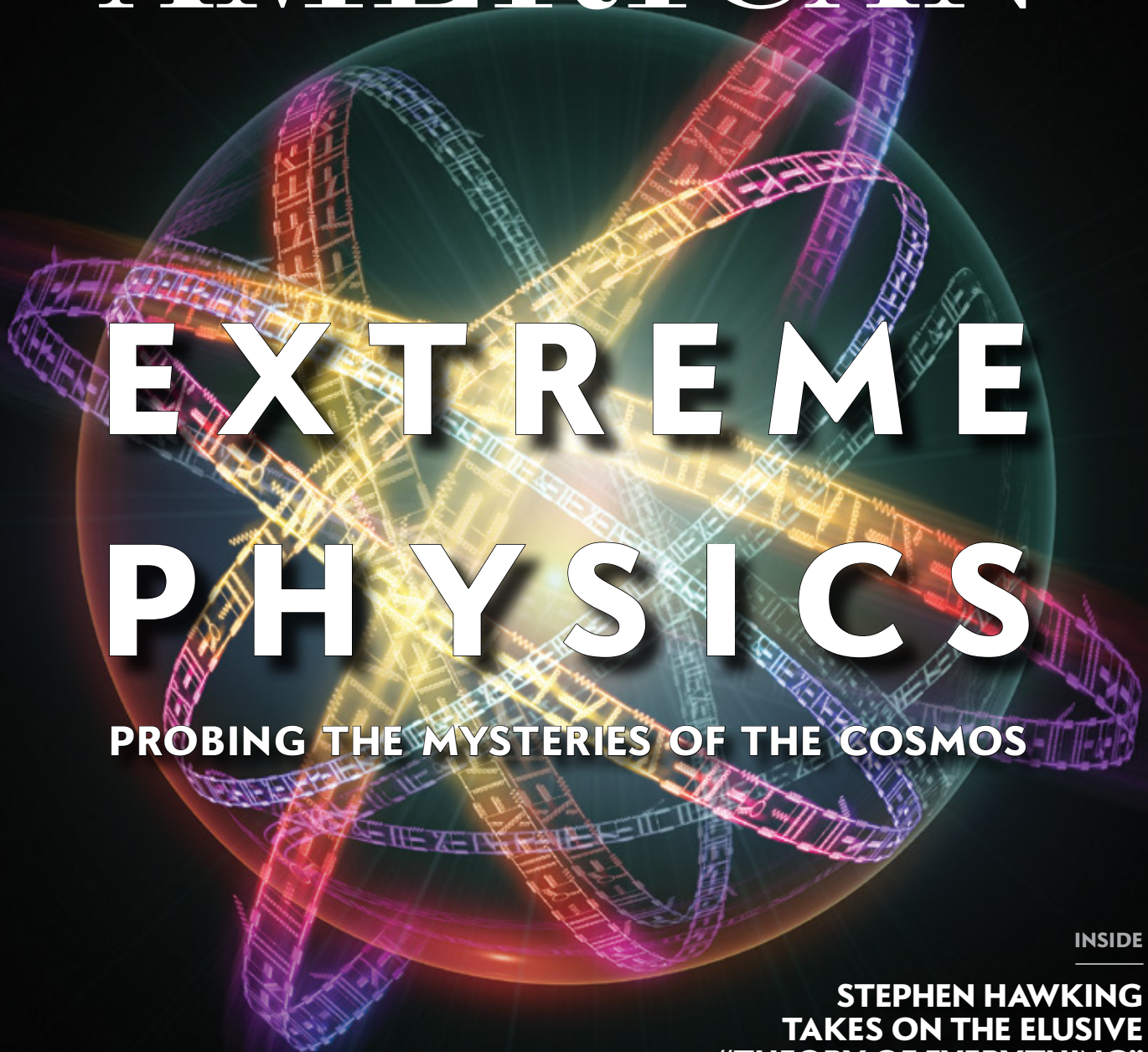
SPECIAL COLLECTOR'S EDITION

SCIENTIFIC AMERICAN



EXTREME PHYSICS

PROBING THE MYSTERIES OF THE COSMOS

INSIDE

STEPHEN HAWKING TAKES ON THE ELUSIVE "THEORY OF EVERYTHING"

WHAT THE HIGGS DISCOVERY REALLY MEANS

HOW QUANTUM THEORY EXPANDS OUR MINDS

$9.99 U.S.

0 72440 38531 3 32>

DISPLAY UNTIL AUGUST 12, 2013

© 2013 Scientific American

EBooks that raise your tablet IQ

Science to enrich your life in a format to fit your lifestyle
From the Editors of Scientific American

Recent years have seen a lot of hype about Higgs. Does it exist? Have scientists really found it? Do we need to worry about it creating a black hole? This eBook chronicles the continuing search for the elusive Higgs boson: the theories, the search and the ongoing questions. In essence, what you need to know to separate Higgs from hype.

A staple of science fiction, the concept of parallel universes allows us to step outside our world and explore questions of "What if?" What if the laws of physics were different? One possibility is that other universes exist where they are. In this eBook, we analyze what science says about the multiverse and the implications for such fundamental questions as, "Are we alone?"

What is time? Time is precious. Time is money. Time is relative. Perhaps one of the reasons we have so many idioms for time is because it is so difficult to define. This eBook summarizes what science has discovered about how time pervades both our physical and our inner mental worlds.

LOOK FOR THESE OTHER TITLES

The Future of Energy: Earth, Wind and Fire
Storm Warnings: Climate Change and Extreme Weather
Beyond Human: A.I. and Genius Machines
Cyber Hacking: Wars in Virtual Space
Forever Young: The Science of Aging

Download Scientific American eBooks now, available wherever eBooks are sold.

| books.scientificamerican.com/sa-ebooks |

SCIENTIFIC AMERICAN™

Apple and iPad are trademarks of Apple Inc., registered in the U.S. and other countries. App Store is a service mark of Apple Inc. Kindle, Kindle Fire, Amazon, the Amazon Kindle logo, and the Kindle Fire logo are trademarks of Amazon.com, Inc. or its affiliates. NOOK is a trademark of Barnes and Noble. ™ Scientific American, Inc.

Infinity's Edge

THE BROADER OUTLINES OF THE PHYSICAL WORLD, FROM QUARKS to the cosmos, have been apparent for decades. Does this mean physicists are about to tie it all up into a neat package? Not at all. Just when you think you know everything, the universe begins to look its strangest.

The detection of what seems for all the world to be a Higgs boson illustrates this idea beautifully. The Large Hadron Collider at CERN near Geneva has provided fodder for physicists from many nations working on a variety of projects but none more important or as eagerly awaited as the confirmation of the Higgs (*page 4*). Our fly-on-the-wall look at scientists poring over data on small perturbations measured with ultrasensitive instruments shows clearly how answers lead to more questions. The data appear to confirm the Standard Model, the theoretical edifice of particle physics, but also reveal subtle and significant ways in which reality diverges from theory, opening up new vistas for exploration.

Just as physicists thought they had found the smallest building blocks of matter—the tiny quarks and leptons—signs have begun to emerge that there are smaller particles still (*page 12*). Proof of their existence would throw physics into complete disarray, however. Further, scientists have never fully understood the neutrino—a ghostly particle that rains down on the planet but usually passes through us, and all matter, unnoticed (*page 20*). The neutrino is therefore another avenue to new inquiries.

On the largest of scales, scientists feel they are closer than ever before to understanding how the universe began and how it will end (*page 36*). Yet other mysteries are as deep as they ever were. Dark energy, which helps to explain why the universe continues to expand at an accelerating rate, remains an enduring puzzle. Now some scientists are beginning to suspect that dark energy does not exist after all. If they are right, basic notions of the universe going back to Copernicus would need a radical revision (*page 58*).

One of the most compelling intellectual problems of our time remains unsolved: How do we tie together the realms of quantum mechanics and general relativity—the very small and the very large? Such a "theory of everything," which has played cat and mouse with scientists for decades, may remain forever out of reach, argue physicists Stephen Hawking and Leonard Mlodinow (*page 90*). If that makes you sad, consider that quantum mechanics, an *Alice in Wonderland* theory that seems to impose severe constraints on what we can know and do, may in fact liberate us by opening a new path to insight (*page 102*). The universe may be finite, but knowledge has not yet been pegged. SA

Fred Guterl
Executive Editor
editors@sciam.com

CORBIS; SCIENTIFIC AMERICAN IS A TRADEMARK OF SCIENTIFIC AMERICAN, INC., USED WITH PERMISSION

SCIENTIFIC AMERICAN™
ESTABLISHED 1845

Extreme Physics is published by the staff of SCIENTIFIC AMERICAN, with project management by:

Editor in Chief and Senior Vice President: **Mariette DiChristina**
Executive Editor: **Fred Guterl**
Issue Editor: **Dawn Stover**
Senior Editor: **Michael Moyer**

Design Director: **Michael Mrak**
Issue Designer: **Lawrence R. Gendron**
Photography Editors: **Monica Bradley, Ann Chin**
Managing Production Editor: **Richard Hunt**
Senior Production Editor: **Michelle Wright**

Copy Director: **Maria-Christina Keller**
Senior Copy Editor: **Daniel C. Schlenoff**
Copy Editor: **Aaron Shattuck**

Editorial Administrator: **Avonelle Wing**
Senior Secretary: **Maya Harty**

Senior Production Manager: **Christina Hippeli**
Advertising Production Manager: **Carl Cherebin**
Prepress and Quality Manager: **Silvia De Santis**
Custom Publishing Manager: **Madelyn Keyes-Milch**
Production Coordinator: **Lisa Headley**

President: **Steven Inchcoombe**
Executive Vice President: **Michael Florek**
Vice President and Associate Publisher, Marketing and Business Development: **Michael Voss**
Director, Integrated Media Sales: **Stan Schmidt**
Vice President, Digital Solutions: **Wendy Elman**
Director, Global Media Solutions: **Jeremy A. Abbate**

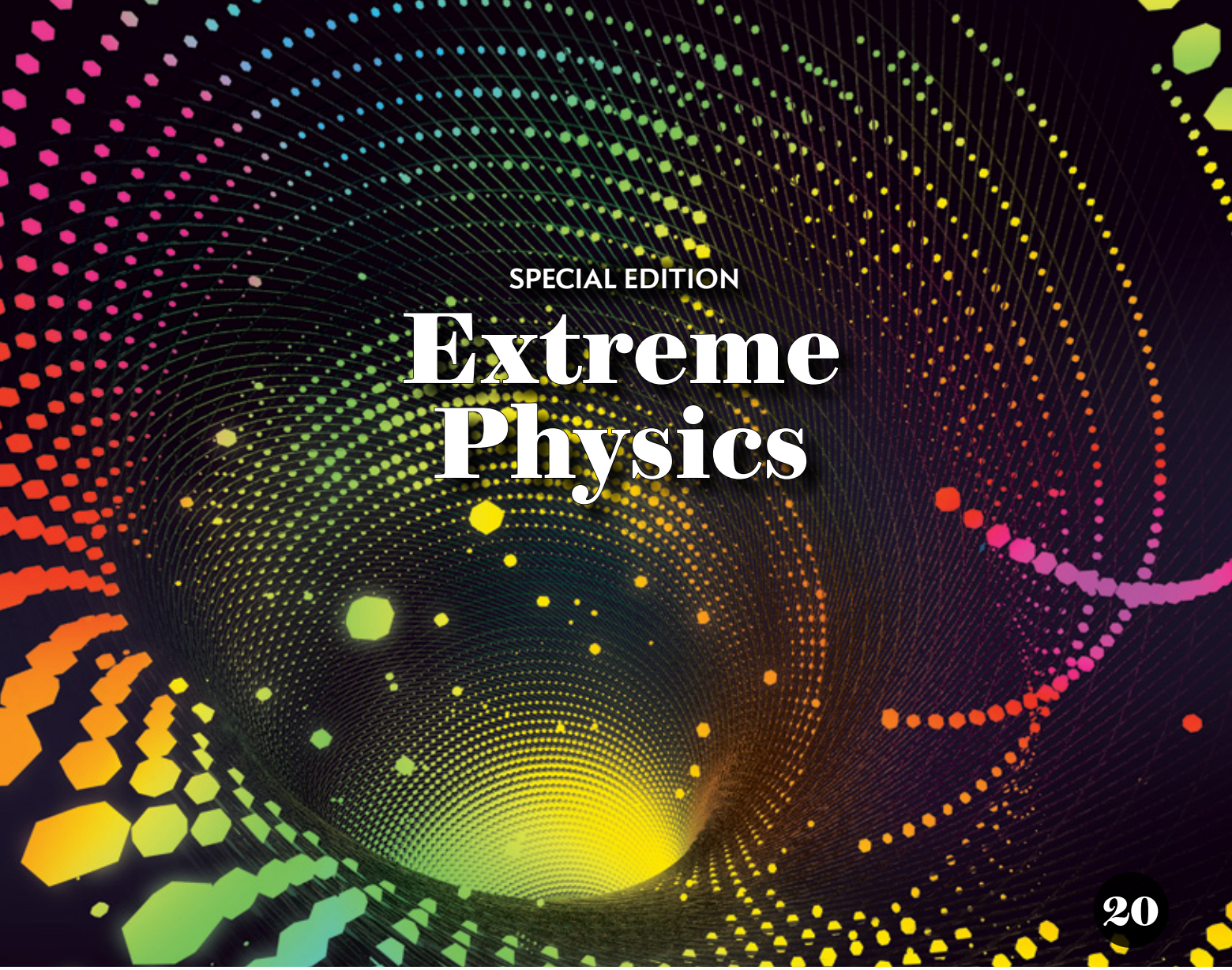
Vice President, Consumer Marketing: **Christian Dorbandt**
Associate Consumer Marketing Director: **Catherine Bussey**
E-Commerce Marketing Manager: **Evelyn Veras**
Senior Marketing Manager/Acquisition: **Patricia Elliott**

Sales Development Manager: **David Tirpack**
Promotion Manager: **Diane Schube**
Promotion Art Director: **Maria Cruz-Lord**
Marketing Research Director: **Rick Simone**
Corporate PR Manager: **Rachel Scheer**
Sales Representative: **Chantel Arroyo**

Director, Ancillary Products: **Diane McGarvey**
Custom Publishing Editor: **Lisa Pallatroni**

Digital Operations Manager: **Scott Rademaker**
Senior Digital Product Manager: **Michael Thomas**
Online Marketing Product Manager: **Zoya Lysak**

© 2013 Scientific American



SPECIAL EDITION

Extreme Physics

20

1 **Introduction**

THE VERY SMALL

4 **The Higgs at Last**
After a three-decade search, scientists appear to have found the elusive particle. Its peculiar properties suggest that a new era in physics is about to dawn.
By Michael Riordan, Guido Tonelli and Sau Lan Wu

12 **The Inner Life of Quarks**
What if the smallest bits of matter actually harbor an undiscovered world of particles? *By Don Lincoln*

20 **Ghostly Beacons of New Physics**
Neutrinos, the strangest beasts in the particle zoo, may soon open the way to unexplored realms.
By Martin Hirsch, Heinrich Päs and Werner Porod

28 **Loops, Trees and the Search for New Physics**
Maybe unifying the forces of nature isn't quite as hard as physicists thought it would be.
By Zvi Bern, Lance J. Dixon and David A. Kosower

THE VERY BIG

36 **Origin of the Universe**
Cosmologists are closing in on the ultimate processes that created and shaped the universe.
By Michael S. Turner

44 **Super Supernovae**
The largest stars die in explosions more powerful than anyone thought possible—some triggered in part by the production of antimatter.
By Avishay Gal-Yam

50 **The Dark Side of the Milky Way**
Dark matter is not just a puzzle. It is a solution.
By Leo Blitz

58 **Does Dark Energy Really Exist?**
Maybe not. The observations that led astronomers to deduce its existence could have another explanation: our galaxy lies at the center of a giant cosmic void.
By Timothy Clifton and Pedro G. Ferreira

COVER IMAGE BY KENN BROWN *Mondolithic Studios*

© 2013 Scientific American

SCIENTIFIC AMERICAN

Volume 22, No. 2, Summer 2013

THE VERY DENSE

66 Portrait of a Black Hole
By adapting a global network of telescopes, astronomers will soon get their first look ever at the dark silhouette of a black hole.
By Avery E. Broderick and Abraham Loeb

74 Naked Singularities
The black hole has a troublesome sibling, the naked singularity. Physicists have long thought—hoped—it could never exist. But could it? *By Pankaj S. Joshi*

82 Black Stars, Not Holes
Quantum effects may prevent true black holes from forming and give rise instead to dense entities called black stars. *By Carlos Barceló, Stefano Liberati, Sebastiano Sonego and Matt Visser*

IN THEORY

90 The (Elusive) Theory of Everything
Physicists have long sought to find one final theory that would unify all of physics. Instead they may have to settle for several.
By Stephen Hawking and Leonard Mlodinow

94 A Quantum Threat to Special Relativity
Entanglement, like many quantum effects, violates some of our deepest intuitions about the world. It may also undermine Einstein's special theory of relativity. *By David Z Albert and Rivka Galchen*

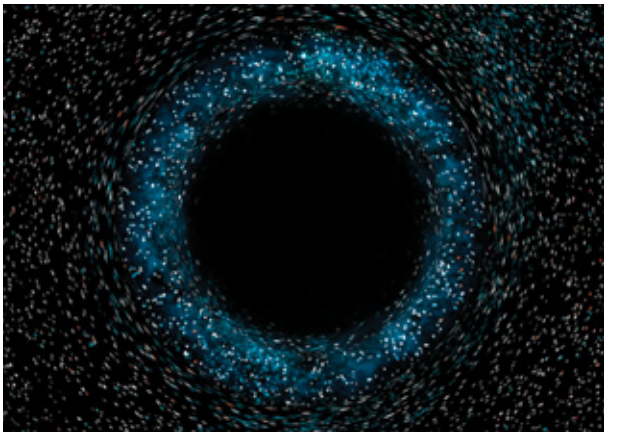
102 Beyond the Quantum Horizon
Once viewed as imposing absolute limits on knowledge and technology, quantum theory is now expanding the power of computers and the vistas of the mind.
By David Deutsch and Artur Ekert

108 Bad Boy of Physics
Leonard Susskind rebelled as a teen and never stopped. Today he insists that reality may forever be beyond reach of our understanding.
Interview by Peter Byrne

112 Signature Analysis
Bubble chamber photographs reveal the telltale tracks of exotic subatomic particles. *By the Editors*

Articles in this special issue are updated from previous issues of *Scientific American*.

4 Higgs hunters have finally bagged a boson.



82 Slow collapse may create a black star.

102 Quantum theory is liberating, not limiting.

EXTREME PHYSICS

Scientific American Special (ISSN 1936-1513), Volume 22, Number 2, Summer 2013, published by Scientific American, a division of Nature America, Inc., 75 Varick Street, 9th Floor, New York, NY 10013-1917. Canadian BN No. 127387652RT; TVQ1218059275 TQ0001. To purchase additional quantities: U.S., $13.95 each; elsewhere, $17.95 each. Send payment to Scientific American Back Issues, 75 Varick Street, 9th Floor, New York, NY 10013-1917. **Inquiries: fax 212-355-0408 or telephone 212-451-8415. Printed in U.S.A.**

Copyright © 2013 Scientific American, a division of Nature America, Inc. All rights reserved.

© 2013 Scientific American

DISCOVERY

THE HIGGS AT LAST

After a three-decade search, scientists appear to have found the elusive particle. Its peculiar properties suggest a new era in physics could be about to dawn

By Michael Riordan, Guido Tonelli and Sau Lan Wu

Illustration by Benjamin Simon and Dirk Schuster

© 2013 Scientific American

© 2013 Scientific American

LATE ON THE EVENING OF JUNE 14, 2012, GROUPS OF GRADUATE students and postdoctoral researchers working on the Large Hadron Collider began peering into a just opened data cache. This huge machine at CERN, the European laboratory for particle physics near Geneva, had been producing tremendous amounts of data in the months since it awoke from its winter-long slumber. But the more than 6,000 physicists who work on the LHC's two largest experiments were wary of unintentionally adding biases to their analysis. They had agreed to remain completely unaware of the results—performing what are called "blind" analyses—until mid-June, when all would suddenly be revealed in a frenzy of nocturnal activity.

Many of the young scientists worked through that night to untangle the newly freed threads of evidence. Although the LHC is a giant collider feeding multiple experiments, only the two largest ones—ATLAS and CMS—had been tasked with finding the Higgs boson, the long-sought particle that would complete the Standard Model of particle physics, the theoretical description of the subatomic world. Each massive detector records the subatomic debris spewing relentlessly from proton collisions in its midst; a detailed, independent accounting of these remnants can reveal fleeting new phenomena, including perhaps the elusive Higgs boson. Yet the detectors have to sift through the particle tracks and energy deposits while enduring a steady siege of low-energy background particles that threaten to swamp potentially interesting signals. It is like drinking from a fire hose while trying to ferret out a few tiny grains of gold with your teeth.

Fortunately, the scientists knew what they were looking for. After the LHC's disastrous start—an electrical splice between two magnets warmed and melted just nine days after the LHC came online in 2008, triggering a powerful spark that punctured the surrounding vessel, released tons of helium and ripped scores of costly superconducting magnets from their mounts—the collider had been collecting reams of data during 2011, enough to pick up an early hint of a Higgs signal.

After that run ended in October for its scheduled winter shutdown, Fabiola Gianotti, then spokesperson for ATLAS, and one of us (Tonelli), then spokesperson for CMS, delivered a special seminar to an overflowing audience in the main CERN auditorium. Both detectors independently found suggestive bumps in the data.

What's more, these telltale hints of a Higgs boson corroborated one another. Both ATLAS and CMS reported several dozen events above the expected background in which two photons came blazing out with combined energies of 125 billion electron volts, or 125 GeV. (GeV is the standard unit of mass and energy in particle physics, about equal to a proton mass.) If proton collisions had created short-lived Higgs bosons, they could have decayed into these photons. Each experiment also found a few surplus events in which four charged leptons (electrons or muons) carried off similar total energies. These could also have been the result of a Higgs [*see box on page 8*]. Such a concurrence of signals was unprecedented. It suggested that something real was beginning to appear in the data.

Yet given the stringent norms of particle physics, none of the signals observed in

IN BRIEF

The Higgs boson, the last missing piece of the Standard Model of particle physics, had for many decades eluded physicists' increasingly elaborate efforts to detect its presence.

Two giant experiments at CERN's Large Hadron Collider had found tantalizing hints of the Higgs in late 2011. At that time, physicists hoped that the spring 2012 run would lead to a discovery.

Physicists hid the new data from even themselves, "blinding" their analyses so as to not introduce bias. In June 2012 they took the first look at the new evidence.

The "Higgs-like" particle that emerged has many of the properties that physicists were looking for. It also held some early surprises that could point the way to the future of physics.

© 2013 Scientific American

2011 were strong enough to allow for claims of a "discovery." Data peaks and bumps like this had often proved ephemeral, mere random fluctuations. And the successful spring 2012 run, which generated more proton collisions in 11 weeks than had come in during all of 2011, could easily have washed out the nascent data peaks, smothering them in background noise.

Of course, the opposite could occur, too. If the bumps were the result of an actual Higgs boson, not just a cruel statistical artifact, all the new data gave researchers a good chance of being able to claim an official discovery—ending this decades-long search and beginning a whole new era in our understanding of matter and the universe.

A THREE-DECADE SEARCH

NEVER JUST ANOTHER PARTICLE, the Higgs boson is the cornerstone of a grand intellectual edifice known as the Standard Model, the interwoven set of theories that constitute modern particle physics. This particle's existence had been suggested in 1964 by Peter W. Higgs of the University of Edinburgh as the result of a subtle mechanism—independently conceived by François Englert and Robert Brout in Brussels plus three theorists in London—that endows elementary particles with mass. The Higgs boson is the physical manifestation of an ethereal fluid (called the Higgs field) that permeates every corner of the cosmos and imbues elementary particles with their distinctive masses. With the discovery of quarks and gluons in the 1970s and the massive, weak-force-bearing *W* and *Z* bosons during the early 1980s, most of the elements of the Standard Model had fallen neatly into place.

Although theorists asserted that the Higgs boson—or something like it—must exist, they could not predict what its mass might be. For this and other reasons, researchers had few clues about where to look for it. An early candidate, weighing in at less than nine times the proton mass, turned up in 1984 at a refurbished, low-energy electron-positron collider in Hamburg, Germany. Yet the evidence withered away after further study.

Most theorists agreed that the Higgs mass should be 10 to 100 times higher. If so, discovering it would require a much larger and more energetic particle collider than even the Fermi National Laboratory's Tevatron, a six-kilometer proton-antiproton collider completed in 1983. That same year CERN began building the billion-dollar Large Electron Positron (LEP) collider, boring a 27-kilometer circular tunnel that crossed the French-Swiss border four times near Geneva. Although LEP had other important physics goals, the Higgs boson was high on its target list.

U.S. particle physicists, encouraged by the Reagan administration to "think big," pushed through grandiose plans for a much larger, multibillion-dollar machine, the Superconducting Super Collider (SSC), in the late 1980s. With a proton-proton collision energy of 40 trillion electron volts (40 TeV, or 40,000 GeV), the SSC was designed to track down the Higgs boson even if it were to come in at a mass near 1,000 GeV.

But after the SSC's projected price tag nearly doubled to $10 billion, Congress voted to kill it in 1993. Dismayed, U.S. Higgs hunters thereafter turned back to Fermilab and CERN to pursue this research. Discoveries and precision measurements made at LEP and the Tevatron soon implied that the Higgs boson should be no more than 200 GeV, which put it potentially within reach of these colliders. In over a decade of searching, however, physicists found no lasting evidence for Higgs-like data bumps.

Hushed conversation suggested something big was building up. Pressure to go public swelled.

During the final LEP runs in the summer of 2000, physicists decided to push the collision energy beyond what the machine was designed to handle. That is when hints of a Higgs boson began appearing. In September two of the four LEP experiments reported evidence for a handful of events with a *Z* boson plus another mystery particle that decayed into two bottom quarks—a particle that looked a lot like a 115-GeV Higgs boson. CERN's then director Luciano Maiani granted the machine a six-week stay of execution that autumn, but during that period researchers could unearth only one more candidate event. It was not sufficient. After a heated debate, Maiani decided to shut LEP down and begin its planned conversion into the LHC, a machine designed to find the Higgs boson.

CLOSING IN ON DISCOVERY

THE LHC IS the most spectacular collection of advanced technology ever assembled. Built inside the original LEP tunnel by hundreds of accelerator physicists and engineers led by project manager Lyndon Evans, it uses little left from that collider. Its principal components include more than 1,200 superconducting dipole magnets—shiny, 15-meter-long cylinders worth nearly $1 million each. Probably the most sophisticated components ever mass-produced, by firms in France, Germany and Italy, they harbor twin beam tubes that are flanked by niobium-titanium magnet coils bathed in liquid helium at 1.9 kelvins, or –271 degrees Celsius. Inside, twin proton beams circulate in both directions at energies up to 7 TeV and velocities approaching light speed.

The beams resemble those of a pulsed laser rather than a flashlight. Each consists of almost 1,400 "bunches," containing up to 150 billion protons apiece—about the number of stars in the Milky Way. Under normal operations, 10 to 30 proton collisions occur during each bunch crossing. That corresponds, however, to around half a billion collisions per second.

Proton collisions are far messier than electron-positron collisions. Theorist Richard Feynman of the California Institute of Technology once compared the process to smashing garbage cans into garbage cans, which means that lots of junk comes out. Protons are composite objects made of quarks and gluons; in the most interesting events, two gluons collide at energies above 100 GeV—and occasionally up to 1 TeV. Physicists, aided by sophisticated detectors, custom-built electronics and state-of-the art computers, try to sift the few events corresponding to interesting physics from the billions of dull, uninteresting ones.

The ATLAS and CMS experiments cannot observe a Higgs boson directly—it would decay into other particles far too quickly. They look for evidence that it was created inside. Depending on the Higgs boson's mass, it could decay into lighter particles in a variety of ways. In 2011 at-

© 2013 Scientific American

tention began to focus on its rare decays into two photons and four charged leptons [*see box below*] because these signals would stand out starkly against the tremendous backgrounds that could easily swamp a Higgs signal.

The year's delay caused by the 2008 magnet disaster gave Fermilab physicists one last shot at making a Higgs discovery. Just before the scheduled Tevatron shutdown in September 2011, the CDF and D-Zero experiments at the collider reported small excesses of events in which bottom quark pairs appeared at combined energies from 125 to 155 GeV. But as in the LEP closure, the researchers could not convince the lab director to grant them a reprieve, and the Tevatron was soon shut down. (In March 2012 these physicists reported a more detailed analysis that showed a bulge centered at 125 GeV, reinforcing the CERN results.)

CROSSING THE LINE

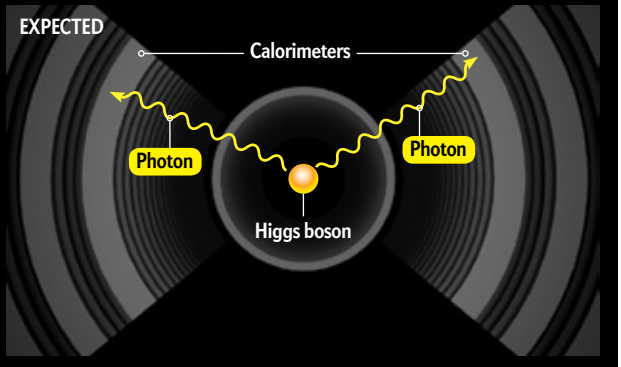
BY MAY 2012 the LHC was producing data 15 times faster than the Tevatron had ever achieved, thanks to efforts of physicists and operators led by accelerator director Stephen Myers. This run was a culmination of two decades of work by thousands of ATLAS and CMS physicists who built and now operate the detectors, designed and now manage a computer system that distributes data around the world, creat-

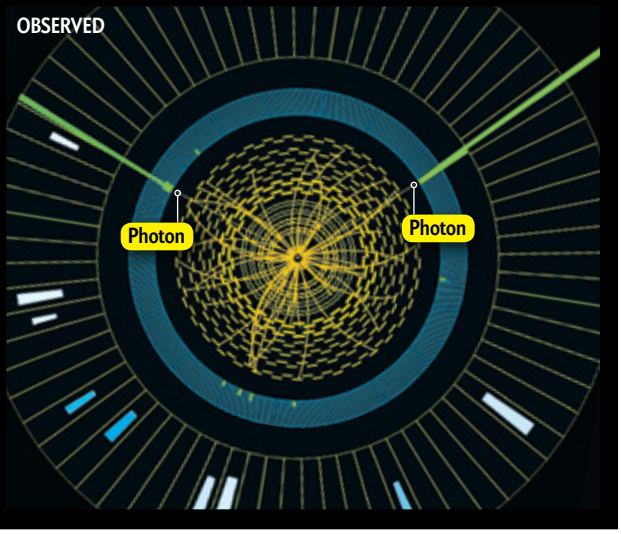
The Delicate, Rare Fingerprints of the Higgs

The Higgs boson is an extremely unstable particle that quickly decays via a number of different processes, or "modes." Unfortunately, many decay modes are indistinguishable from the thunderous din of ordinary background events that result from 500 million proton-proton collisions every second. The ATLAS and CMS experiments are designed to spot the occasional interesting

Photons

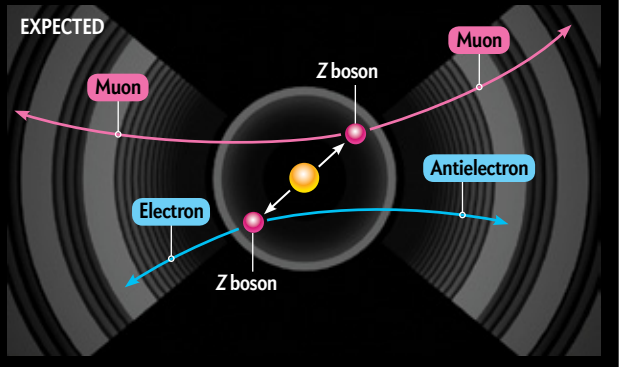
Each detector includes multiple calorimeters, devices for measuring the energy of particles. The innermost calorimeter is particularly alert for photons. These are absorbed in the calorimeter and create tiny electrical signals. If a Higgs decays into two photons, the detector can measure their total energy at extremely high accuracy, which helps to precisely reconstruct the mass of the newly found particle.

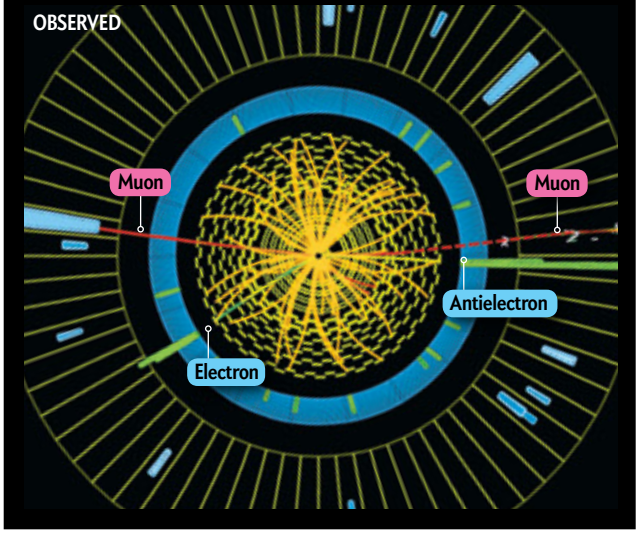




Z Bosons

The Higgs may decay into a pair of *Z* bosons, each of which can decay into an electron paired with an oppositely charged antielectron or two muons. An inner tracker and calorimeter measure the electrons, while muons fly out, leaving footprintlike tracks as they go. High magnetic fields bend the path of electrons and muons during their trip, allowing for a high-resolution measurement of their energy and the original Higgs mass.





COURTESY OF CERN (*observed signals from CMS detector*)

Illustration by George Retseck

© 2013 Scientific American

ed novel hardware and computer software to identify the most interesting collisions, and wrote the algorithms that dig out the most pertinent events from the great morass of data being recorded. They all worked feverishly, anticipating a discovery. So when the researchers opened their data sets in mid-June, they had torrents of events to sift through. After graduate students and postdocs worked through the night, they anxiously prepared to reveal what had turned up.

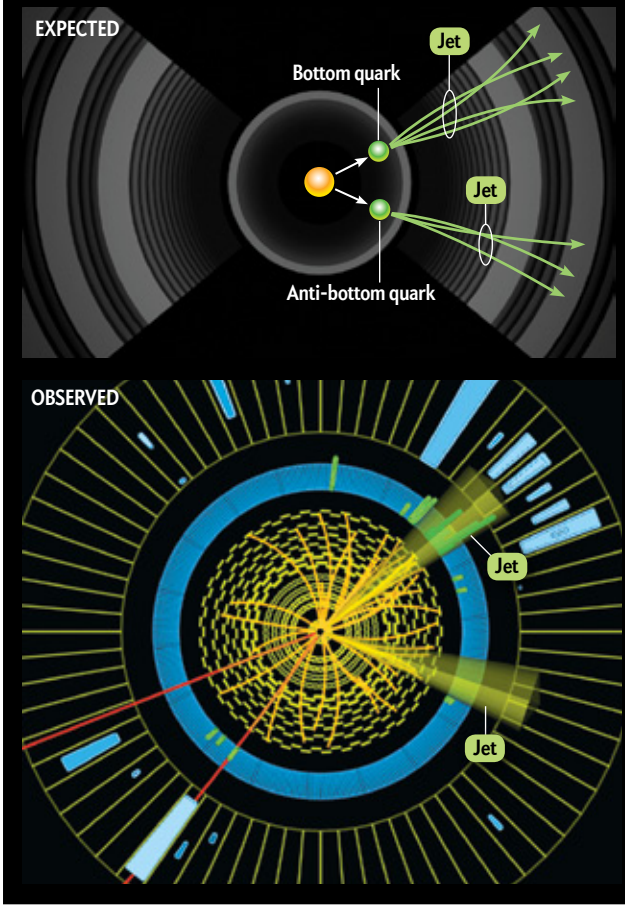
It was a hot afternoon on June 15, 2012, when CMS physicists began gathering in Room 222 of the CERN filtration plant to hear the young physicists' reports. Soon the room was crowded with hundreds of collaboration members—out of about 3,000 in all—many of them standing or sitting on the floor. Few had slept much the night before. Tension and excitement gripped the room.

The first speaker discussed one possible Higgs decay route, or "channel," into pairs of *W* bosons. A small excess of events appeared in the low-mass region of most interest, but the faint signal generated no great excitement. Then presentations on the rare four-lepton and two-photon decays came one after the other. Now it indeed looked like a Higgs boson was showing up at long last. The signals from the 2012 data were occurring again

events that might come from the Higgs decay and throw much of the rest away. The drawings below show four of the most important decay modes that experiments use to search for the Higgs, along with images of actual Higgs-like signals that CMS observed in the 2011 and 2012 runs. (Because the discovery is statistical in nature, no single event can be used as definitive proof.)

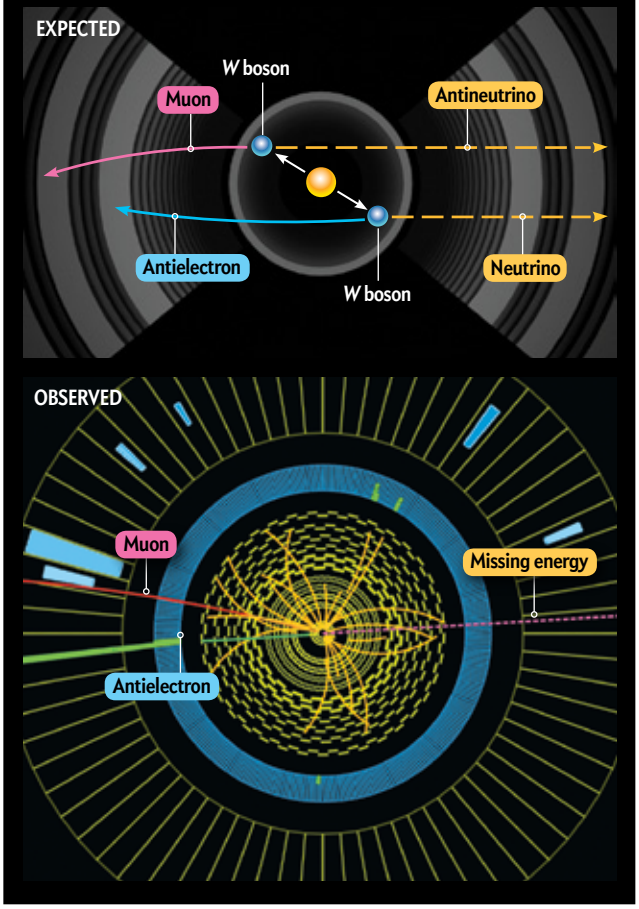
Bottom Quarks

The Higgs can also decay to a bottom quark and its antiparticle, each of which decays into a tight "jet" of secondary particles called hadrons (composite particles made of quarks). These hadrons fly through the detector's inner layers and deposit their energy in the outer calorimeters. Unfortunately, many ordinary collisions also generate jets of hadrons from bottom quarks, which makes it difficult to separate these Higgs events out from the background.



W Bosons

The Higgs can also decay to two *W* bosons, each of which can decay into an electron, antielectron or muon, plus a neutrino or antineutrino. Neutrinos are nearly impossible to detect—they fly out of the detector as if they were never there, taking with them some of the event's energy. Researchers use this missing energy to infer their presence, but the missing energy also prevents them from accurately reconstructing the mass of the original Higgs boson.



© 2013 Scientific American

TIMELINE

Five Decades of the Higgs

Last summer's discovery of a Higgs-like particle marks the culmination of a decades-long search. In the years before the Standard Model of particle physics came together, researchers realized that they had no explanation for why particles should have mass. A series of theoretical insights suggested that a new type of field—now called the Higgs field—could slow particles down and give them their inertia. This field should have a particle counterpart, and so the search for the Higgs was on.

August 1964
THE PAPERS
François Englert and Robert Brout publish the first of three papers proposing a particle and mechanism that will come to be named after Peter W. Higgs (*left*), author of the second paper, which is published two weeks later. Gerald Guralnik, Carl Hagen and Tom Kibble publish the third paper in November.

January 1983
***W* BOSON DISCOVERED**
One of the last missing pieces of the Standard Model is uncovered when an experiment at the Super Proton Synchrotron at CERN near Geneva spots *W* bosons for the first time.

August 1979
GLUON DISCOVERED
Scientists first observe the gluon, the particle responsible for nuclear forces, at the DESY laboratory in Hamburg, Germany. Theorists calculate that gluon fusion will create more Higgs bosons than any other process.

July 1989
NEW COLLIDER COMES ONLINE
In an effort to bag bigger quarry, CERN constructs the Large Electron Positron (LEP) collider inside a circular, 27-kilometer-long tunnel.

in the same vicinity—near 125 GeV—that had so tantalized researchers six months earlier. Scientists realized almost immediately that if they were to combine the new data with the 2011 results, chances were good that CMS could claim a Higgs discovery. The crowd cheered at the end of the two key presentations.

Similar revelations occurred in the ATLAS experiment. Spontaneous celebrations broke out in several groups when they first glimpsed the new data. Yet it took more than a week of long workdays and sleepless nights before ATLAS physicists were certain that they had enough good events to conclude that the chances that their results were due to random fluctuations were less than one in three million—corresponding to the stringent "five sigma" standard that particle physicists require before claiming a discovery. At the thrilling moment of recognition, one ATLAS group of about a dozen physicists, meeting in Building 32 on the afternoon of June 25, 2012, erupted in loud clapping and cries of joy, which echoed down the hallway.

By that time word of a discovery had leaked out. Worldwide interest began growing so intense that secrecy was placed at a premium. There were to be no further leaks before the official word was presented, particularly because the exact content of documents under preparation could change. ATLAS members were not supposed to talk about the recent results with CMS physicists, nor vice versa. Individual physicists, however, could not resist discussing the news many had awaited so long. Hushed conversations in the CERN cafeteria and corridors suggested that something big was building up. Pressure to go public swelled.

CERN director Rolf-Dieter Heuer got an early glance at the findings in a June 22, 2012, meeting with Gianotti and Joseph Incandela of the University of California, Santa Barbara, Tonelli's successor as CMS spokesperson. Heuer decided that the evidence was strong enough to make public. He immediately informed the CERN Council (its governing body) to keep them abreast of the fast-moving developments. Heuer then decided to hold a joint seminar at CERN on July 4, 2012, timed to coincide with the opening of the 36th International Conference on High Energy Physics in Melbourne, Australia, followed by a CERN press conference.

The night before the seminar, hundreds of physicists dozed fitfully in the hallways outside the locked main auditorium, desperately hoping to get one of the unreserved seats remaining inside. Myers, Evans and four prior CERN directors who had been heavily involved with the LHC since its conception were seated in the front row. Having just flown to Geneva, Peter Higgs walked in to warm, sustained applause and sat down next to Englert.

Incandela and then Gianotti showed blizzards of slides about the new data and results, mostly covering the 2012 measurements. As in December, graphs of two-photon data revealed striking peaks jutting out at 125 to 126 GeV. And this time around, the experiments had more than a dozen extra events in which a heavy particle had exploded into four charged leptons at 125 GeV. Subtle peaks had begun to form in that channel, too.

That clinched it. Combining this result with the two-photon one, CMS and ATLAS independently concluded that the chances that the apparition was a fluke, due to random fluctuations, were less than one in three million. It had to be real. When the camera panned to Higgs, he could be seen pulling out a handkerchief to wipe his eyes.

"I think we have it," exulted Heuer, wrapping up the seminar to sustained applause. "We have a discovery," he went on, guardedly using the word at last. "We have observed a new particle consistent with a Higgs boson."

A NEW ERA IN PHYSICS?

FEW PHYSICISTS DOUBT that a heavy new particle has turned up at CERN, but there is still debate about its exact nature. CERN officials initially spoke cautiously on this

COURTESY OF CERN (*facilities and collision events*); AP PHOTO (*Higgs*)

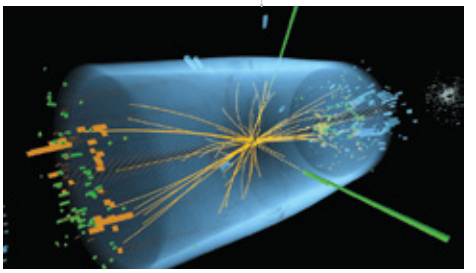
© 2013 Scientific American

November 2, 2000
THE END OF AN ERA
The LEP collider closes so that construction may begin on CERN's Large Hadron Collider (LHC), the machine that will eventually find the Higgs.

September 10, 2008
ALL SYSTEMS GO
The first proton beams shoot around the newly finished LHC.

September 19, 2008
DISASTER STRIKES
After an electrical splice between two magnets warms and melts, a powerful spark punctures a magnet vessel and releases tons of liquid helium. More than 50 magnets rip from their mounts or are otherwise damaged.

September 2000
LAST PUSH FOR THE HIGGS
Scientists at LEP detect hints of the Higgs boson just as the machine is scheduled to be permanently shut down. Administrators offer a six-week reprieve and push the machine past its design energy but for naught. We now know the weak signal was not the Higgs after all: it was at the wrong mass.



July 4, 2012
HIGGS-LIKE PARTICLE FOUND
CERN scientists announce that they have discovered a Higgs-like particle at 125 GeV.

—Compiled by Marissa Fessenden

question, calling it a "Higgs-like" boson, but in March announced that it is indeed a Higgs—though not necessarily the only one. Although physicists have not yet proved beyond a doubt that the new particle has the required property of zero "spin," preliminary data strongly favor that value. The ATLAS experiment continues to observe more two-photon events than expected, while CMS is reporting results consistent with Standard Model expectations based on a similar amount of data. Could something be amiss here?

Since July 2012 attention has become focused on whether the new particle is indeed "the" Higgs boson as predicted by the Standard Model. That question can be resolved by determining how the new boson decays into other particles. Results released by ATLAS and CMS since July 2012 show that the Higgs signal has greatly improved, while difficult-to-measure decays into bottom quarks and tau leptons are beginning to appear at about the expected frequency. Meanwhile Fermilab physicists published evidence from the Tevatron for decays of the new particle into bottom quarks. Analyses of the combined LHC and Tevatron data by CERN theorists John Ellis and Tevong You indicated that the new particle, as they put it, "does indeed walk and quack very much like a Higgs boson."

The new particle's connection with a pair of high-energy photons has stimulated intrigue. Because the Higgs field imbues elementary particles with mass, it should interact more strongly with heavier particles. Photons have no mass, so the Higgs boson produces them via a mechanism involving other, massive particles. Additional heavy particles (which are required by supersymmetry and other theories) could enhance the process—as may be happening, based on early data. If the tendency holds up, it will strongly suggest physics beyond that described by the Standard Model.

The Higgs discovery marks the end of a long era in particle physics and the beginning of an exciting new phase. After decades in the doldrums, the discipline is energized once again by the heady intercourse of theory and experiment. Questions abound that may find answers from further research on this fascinating particle or its potential partners. Does it play a role in the inflation mechanism considered the force driving the big bang origins of the universe? Does it interact with dark matter particles thought to inhabit the cosmos? And what higher-energy mechanism or process, if any, shields the fragile vacuum from instabilities that may threaten the existence of the universe as we know it?

Although we celebrate the triumph of the Standard Model, such a lightweight Higgs boson should be extremely sensitive to physics lying beyond it. The particle opens up a fabulous new laboratory for further experimentation. Are its properties exactly as predicted? The apparent discrepancies in the early data could be random fluctuations that disappear in months to come. Or perhaps they are offering subtle hints of intriguing new physics. SA

Michael Riordan is a historian of science and the author of numerous books, including *The Hunting of the Quark* (Simon & Schuster, 1987). He is writing a history of the ill-fated Superconducting Super Collider.
Guido Tonelli is a professor at the University of Pisa in Italy and a researcher at Italy's National Institute of Nuclear Physics. He has been working on the CMS experiment at CERN since 1993 and served as its spokesperson in 2010 and 2011.
Sau Lan Wu has been searching for the Higgs boson for more than 20 years, first at the LEP collider and since 1993 as a member of the ATLAS experiment at CERN. She is Enrico Fermi Distinguished Professor of Physics at the University of Wisconsin–Madison.

MORE TO EXPLORE

The Mysteries of Mass. Gordon Kane in *Scientific American*, Vol. 293, No. 1, pages 40–48; July 2005.
The Coming Revolutions in Particle Physics. Chris Quigg in *Scientific American*, Vol. 298, No. 2, pages 46–53; February 2008.
The Discovery Machine. Graham P. Collins in *Scientific American*, Vol. 298, No. 2, pages 39–45; February 2008.
The ATLAS experiment: **http://atlas.ch**
The CMS experiment: **http://cms.cern.ch**

© 2013 Scientific American

SUBATOMIC STRUCTURES

THE INNER LIFE OF QUARKS

What if the smallest bits of matter actually harbor an undiscovered world of particles?

By Don Lincoln

© 2013 Scientific American

Photograph by Craig Cutler

© 2013 Scientific American

THE UNIVERSE IS A COMPLEX AND INTRICATE PLACE. WE CAN move easily through air and yet not through a wall. The sun transmutes one element to another, bathing our planet in warmth and light. Radio waves have carried a man's voice to Earth from the surface of the moon, whereas gamma rays can inflict fatal damage on our DNA. On the face of it, these disparate phenomena have nothing to do with one another, but physicists have uncovered a handful of principles that fuse into a theory of sublime simplicity to explain all this and much more. This theory is called the Standard Model of particle physics, and it encapsulates the electromagnetic forces that make a wall feel solid, the nuclear forces that govern the sun's power plant, and the diverse family of light waves that both make modern communications possible and threaten our well-being.

The Standard Model is one of the most strikingly successful theories ever devised. In essence, it postulates that two classes of indivisible matter particles exist: quarks and leptons. Quarks of various kinds compose protons and neutrons, and the most familiar lepton is the electron. The right mix of quarks and leptons can make up any atom and, by extension, any of the different types of matter in the universe. These constituents of matter are bound together by four forces—two familiar ones, gravity and electromagnetism, and the less familiar strong and weak nuclear forces. The exchange of one or more particles known as bosons mediates the latter three forces, but all attempts to treat gravity in the microrealm have failed.

The Standard Model leaves other questions unanswered as well, such as: Why do we have four forces and not some other number? And why are there two types of fundamental particles rather than just a single one that handles everything?

These are intriguing problems. Nevertheless, for a long time now a different puzzle has captured my attention and that of many other physicists. The Standard Model views quarks and leptons as indivisible. Astoundingly, though, various clues imply that they are instead built of still smaller components. If quarks and leptons are not fundamental at all, and smaller bits do in fact exist, their presence will force extensive revisions of our theories. Just as nuclear power was inconceivable before Ernest Rutherford discovered the structure of the atom in 1911, unveiling another layer of the subatomic onion will certainly reveal phenomena we cannot yet imagine.

GENERATION GAPS

RESOLVING THIS ISSUE requires scientists to smash particles together at extremely high energies. Since the observation of quarks in the 1970s, we have lacked the tools that might allow us to peer inside them. But now the Large Hadron Collider (LHC) at CERN near Geneva—the same machine that recently confirmed the existence of a Higgs boson, the last undocumented particle in the Standard Model—is gaining speed and could be up to the task.

The first hints of structure in quarks

IN BRIEF

In 1869 Dmitri Mendeleev created the periodic table by noticing that properties of the chemical elements fit into a repeating pattern, which physicists later explained as a consequence of atomic structure. A similar story may be playing out in particle physics today. **The 12 known** elementary particles have their own repeating patterns, suggesting that they are not truly fundamental but actually tiny balls containing smaller particles, which physicists tentatively call preons. **Other evidence** argues against this possibility. The Large Hadron Collider at CERN, along with several lesser-known experiments, may finally settle the question.

© 2013 Scientific American

and leptons emerged from research into another—still unsolved—poser, related to the numbers of different kinds of quarks and leptons that have been discovered. Protons and neutrons consist of two types of quarks, called the up quark and the down quark. Up quarks have +2/3 the electrical charge of the proton, and down quarks have –1/3 of the proton's charge. Although only these two types of quarks, plus electrons, suffice to make up the matter of the universe, other quarks have been observed. The strange quark has the same charge as the down quark, but it is heavier. The bottom quark is an even heavier version. Similarly, the charm quark is a heavier cousin of the up quark, with the superheavy top quark rounding out the up quark family. Particle physicists have observed all these quarks, but the four heavier ones decay, in fractions of a second, into the lightest two.

The electron also has heavy, unstable cousins, the muon and the even heavier tau lepton, both of which have the same charge as the electron. And the known menagerie of particles includes three copies of neutrinos, all of which are superlightweight and electrically neutral.

This cornucopia naturally led physicists to ask: Given that the up quark, down quark and electron are the only particles necessary to build a universe, why do they have so many cousins? The question can be encapsulated in Nobel Prize–winning physicist I. I. Rabi's oft-quoted quip when he learned of the discovery of the muon: "Who ordered that?"

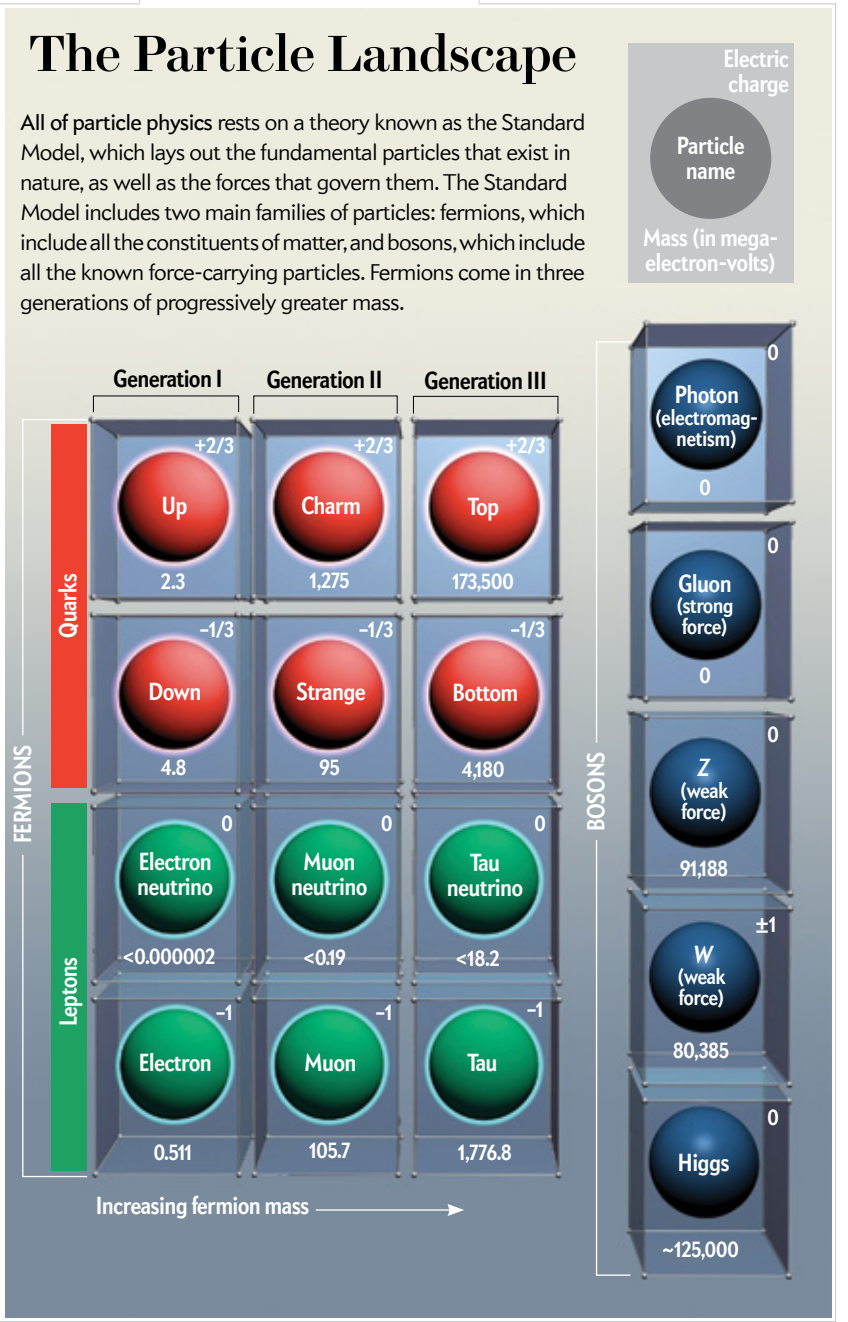
One way scientists went about tackling the mystery of populous particle families was to construct a chart delineating the features of all known elementary particles [*see box at right*], analogous to the periodic table of chemical elements. The periodic table offered physicists the first hints that the chemical elements might not be fundamental, that systematic patterns in the atom's inner structure might account for similar properties of elements in particular rows and columns.

The table of quarks and leptons has three columns called generations (which is why the mystery of particle multiplicity is now referred to as the generation problem). Generation I, at the far left, includes the up and down quark as well as the electron and electron neutrino—everything needed to explain our familiar universe. Generation II contains the somewhat more massive versions of the same particles; generation III has the most massive of all.

The Standard Model treats the quarks and leptons as pointlike particles without any internal structure. But the patterns within the table, as within chemistry's periodic table, raise the possibility that the differences in generations stem from the configuration of even smaller building blocks of matter within quarks and leptons.

Another historical precedent, near the dawn of the 20th century, that may have relevance in the search for the quark's underlying structure is the discovery of radioactive decay. Through a process not understood at the time, one element can transmute into another. We now know that by changing the number of protons and neutrons in the nucleus, it is possible to achieve the goal of medi-

THE STANDARD MODEL



Illustrations by Malcolm Godwin

© 2013 Scientific American

eval alchemists and convert lead into gold. The range of possible transmutations is even wider, as nuclear alchemy can even convert a neutron into a proton (or the reverse) by changing the identity of their constituent quarks. This transformation occurs via the weak nuclear force, which can also transmute leptons, although quarks cannot be changed into leptons, or vice versa. Just as the conversion of one element into another reflects the complex inner workings of the atom, so the metamorphosis of the quarks and leptons may provide yet another hint of even finer details within those particles.

PART AND PARTICLE

MANY HYPOTHETICAL building blocks for quarks and leptons have emerged, each with a different name, but the term "preon" has stuck as a generic descriptor for all of them. In most cases, the same name applies to the constituents of the particles that carry the forces acting on these bits of matter.

As an illustration, consider a straightforward model proposed independently in 1979 by Haim Harari, then at the Stanford Linear Accelerator Center, and Michael A. Shupe, then at the University of Illinois at Urbana-Champaign, and subsequently extended in 1981 by Harari and his student Nathan Seiberg, both then at the Weizmann Institute of Science in Rehovot, Israel [*see box at left*]. They proposed that two kinds of preons exist, one with an electrical charge of +1/3 and one with a charge of zero; in addition, each of these preons has an antimatter companion with opposite charge: –1/3 and zero, respectively. These preons are fermions—particles of matter—and each quark and lepton contains a unique mix of three preons. Two preons of +1/3 charge and one with zero charge, for instance, make an up quark, whereas the up quark's antimatter counterpart contains two preons of –1/3 charge and one with zero. The force-carrying bosons, meanwhile, consist of unique six-preon combinations. The positively charged *W* boson, for instance, which carries the weak nuclear force that acts on both quarks and leptons, has three +1/3 preons and three zeroes.

Using a series of sensible assumptions, Harari and Shupe postulated the preon content of all the particles of the first generation. The same building blocks can also account for gluons, the subatomic particles that mediate the strong nuclear force to bind quarks together inside protons and neutrons, as well as the other force-carrying bosons.

The trick in formulating any underlying structure of the well-known quarks, leptons and bosons is accounting for the myriad interactions of those particles and forces. Indeed, preons can provide a sensible language for describing subatomic processes. For instance, consider an up quark colliding with an antimatter down quark, yielding a positively charged *W* boson that decays into an antimatter electron, or positron, and an electron neutrino. In the preon model devised by Harari and Shupe, the incoming quarks, with their three preons apiece, combine in the collision to generate a *W* boson, now containing all three +1/3 charges and all three zero charges. Then the *W* boson comes apart, spitting out a different configuration of the same six preons: one positron (with the three +1/3 charg-

PREONS FOR BEGINNERS

A Particle Cookbook

Physicists have proposed various concepts for preons—the particles that might make up quarks and other elementary particles. One notable model (*below*) was proposed in 1979 by Haim Harari, then at the Stanford Linear Accelerator Center, and Michael A. Shupe, then at the University of Illinois at Urbana-Champaign. Their scheme posits two kinds of preons and an antimatter version of each, which could comprise both the particles of matter, or fermions (*top*), and the particles of force, or bosons (*bottom*).

Particles of Matter (Fermions)

The two preons in this model can be represented as + and 0. The + has a charge of +1/3, and the 0 has a charge of zero. Each has a corresponding antimatter companion with the opposite charge: – (–1/3) and $\bar{0}$ (zero). In Harari and Shupe's model, the quarks and leptons are each composed of three preons.

CHARGE	PREON CONTENT	PARTICLE
+1	+ + +	Antimatter electron
+2/3	+ + 0	Up quark
+1/3	+ 0 0	Antimatter down quark
0	0 0 0	Electron neutrino
0	$\bar{0}\,\bar{0}\,\bar{0}$	Electron antineutrino
–1/3	– $\bar{0}\,\bar{0}$	Down quark
–2/3	– – $\bar{0}$	Antimatter up quark
–1	– – –	Electron

Particles of Force (Bosons)

In groupings of twos and sixes, the same preons can also describe the bosons, whose transfer mediates the subatomic forces: electromagnetism (photon), the strong nuclear force (gluon) and the weak nuclear force (W^+, W^-, *Z* bosons). The details for gluons, the subatomic particles that bind quarks together inside atomic nuclei, are a bit more complicated and are omitted here.

CHARGE	PREON CONTENT	PARTICLE
+1	+ + + 0 0 0	Positive *W* boson
–1	– – – $\bar{0}\,\bar{0}\,\bar{0}$	Negative *W* boson
0	0 0 0 $\bar{0}\,\bar{0}\,\bar{0}$ + + + – – – + + – – 0 $\bar{0}$ + – 0 $\bar{0}\,\bar{0}\,\bar{0}$	*Z* boson (four versions)
0	+ –	Photon

© 2013 Scientific American

es) and one electron neutrino (with the three zero charges).

Thus far I have discussed what might be called quark and lepton numerology. This is just a counting game, like balancing chemical or math equations, albeit a serious and feasible one. To be successful, a preon model must explain the quarks and leptons with a small number of building blocks and a few governing rules. After all, the hope is to find an underlying order that unifies superficially different particles, not a system of ad hoc definitions that accounts for their properties on a case-by-case basis. Such an explanation has been accomplished with preons, both in the Harari-Shupe model and in its successful competitors.

You may have noticed, however, that the discussion so far has included only the first generation of quarks and leptons. Things get murkier when we turn our attention to the second and third generations. Within the model proposed by Harari and Shupe, the higher generations are hypothesized to be excited states of the first-generation configurations. Just as electrons jump from one energy level to another in atoms, some unknown mechanism is thought to bind the preons together in a way that allows for multiple-particle generations from the same ingredients.

If this explanation seems a bit like hand waving, it is. Many of the details have not been worked out. The theoretical studies that first proposed the idea of quarks had a similar level of sophistication. It was only later that the strong force, which binds quarks together into protons and neutrons, was described mathematically. Still, the generation problem remains conspicuously unexplained, so several physicists have proposed competing models, including one in which one of the preons carries the generation number as well as a new charge called hypercolor, which binds the preons inside the quarks and leptons.

Although I have described a single theory of preons, do not be misled into thinking it is the only one out there. My theoretical colleagues are very smart and very creative. Literally hundreds of papers have been written proposing other preon models, although these models are often variations on a small number of basic themes. Some have preons with 1/6 charge, rather than the 1/3 in Harari and Shupe's model. Others have five preons in the quarks and leptons as opposed to three. Still others propose a mix of fermion preons and boson preons or different preon contents for the bosons than those laid out in the bottom table on the opposite page. The possibilities are actually quite rich. We physicists need more data to help weed out the alternatives.

Beyond inherent fascination with the notion that the smallest known pieces of matter might have smaller pieces still, many of us are interested in preons for another reason. If they exist, they could have something profound to say about another outstanding mystery in particle physics. The Standard Model postulates that the Higgs field is the source of mass for fundamental particles. Massive particles feel a sort of drag as they move through this ubiquitous field, whereas massless particles such as the photon glide through unmolested. If the preons that make up the second and third generations are the same as the first, presumably something about the preons enables the higher-generation particles to interact more with the Higgs field than the first generation does, thus giving the higher generations their greater mass. Whereas the Higgs mechanism can account for the masses of the particles, it cannot predict them.

Just as nuclear power was inconceivable before the discovery of atomic structure, unveiling a new layer of matter would reveal phenomena we cannot imagine.

Until a deeper theory is invented, the mass of subatomic particles can be determined only by measuring them one by one. Presumably by understanding the structure of quarks and leptons and how the generations differ, we will learn much more about the Higgs mechanism.

POTHOLES AND DETOURS

I SHOULD NOTE THAT preon theory is not without its problems. For starters, all experimental efforts to see preons have failed. That failure is disappointing but could stem simply from having inadequate equipment. Experimental questions aside, some concerns are intrinsic to the theory itself. It is a natural feature of "confining theories," so called in this context because the preons are confined inside the quarks and leptons, that the relevant masses are inversely proportional to the confinement size. Because quarks and leptons are much smaller than protons, this rule implies that a quark made of confined preons would be much more massive than a proton, which is itself made of quarks. The proton whole would be less than the sum of its parts—less, indeed, than the individual parts themselves.

Although this problem may seem insurmountable, physicists have managed to get around a similar kink related to bosons. A quark and antiquark, for instance, can make up a boson called a pi meson, in which the confinement conundrum also seems to pose a problem. Using an idea sketched out in 1961 by Jeffrey Goldstone, then at CERN, however, theorists have long realized that symmetries in the underlying theory could overcome this difficulty. Thus, the lightness of the pi meson was not really a surprise. Unfortunately, this approach applies only to bosons, not to fermions such as quarks. Yet in 1979 Gerard 't Hooft of Utrecht University in the Netherlands worked out a related approach that does work for fermions. Whether 't Hooft's concept is borne out in actual particles remains unclear, but his ideas have at least shown that the theoretical roadblock of quark masses is not as formidable as it first appeared.

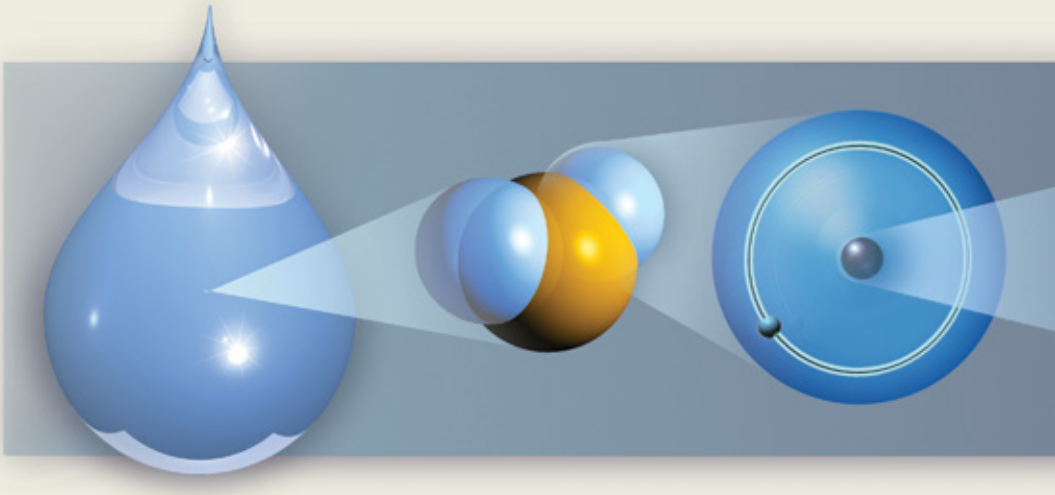
Preons are not the only avenue physicists have explored in hopes of solving the generation problem. One prominent alternative is the idea of superstrings, in which the ultimate building blocks of matter are not subatomic particles but tiny vibrating strings. Metaphorically, each of the Standard Model particles can be thought of as strings playing a different note and all of reality as an orchestra of superstrings playing a grand cosmic symphony. Happily, preons and superstrings can amiably coexist because the size scale of superstrings is much smaller than that of quarks and leptons. If superstrings exist, they could well make up not quarks and leptons but rather pre-

© 2013 Scientific American

SCALES OF SIZE

Zooming In

If preons exist, they are almost unimaginably tiny. They would have to fit inside the quark, currently the smallest known particle of matter, which itself must be small enough to fit into the proton. In fact, all experiments to date are consistent with the quark having a size of zero, which would preclude any inner structure. But future experiments will zoom in for a closer look. A nonzero size for the quark would give the preon hypothesis a major boost.



Raindrop
10^{-3} meter

Water molecule
2.8×10^{-10} m

Hydrogen atom
1.1×10^{-10} m

ons or even pre-preons or pre-pre-preons, depending on how many undiscovered onion layers of matter exist.

Another alternative to the idea of preons as ordinary, albeit undiscovered particles emerged in 2005, when Sundance Bilson-Thompson of the University of Adelaide in South Australia devised a way of describing preons as twisted braids of spacetime. This model is still in its infancy, but physicists are exploring its implications, not least because it offers one possible path to integrating a long-sought quantum theory of gravity into the Standard Model.

PROOF IN THE PREON PUDDING

PHYSICS IS ULTIMATELY an experimental science. No matter how clever the theory, if it fails to agree with measurement, it is wrong. So what can experimentalists do to prove or disprove the existence of preons? The Standard Model successfully describes the quarks, leptons and bosons of the universe without invoking preons, so physicists must look for subtle deviations from the Standard Model's predictions—tiny cracks in the edifice of modern physics. Two facets of the model in particular look like attractive areas to explore more closely.

The first is size. The Standard Model treats the quarks and leptons as pointlike—that is, particles with zero size and no inner structure. Finding a nonzero size for those particles would provide powerful evidence for preons. Measurements have shown that protons and neutrons have a radius of about 10^{-15} meter. Experiments at the world's leading particle colliders, past and present, have searched for evidence that quarks or leptons also have a measurable size. Thus far all the data are perfectly consistent with zero size or with a nonzero size as small as about 0.0002 to 0.001 times the size of a proton.

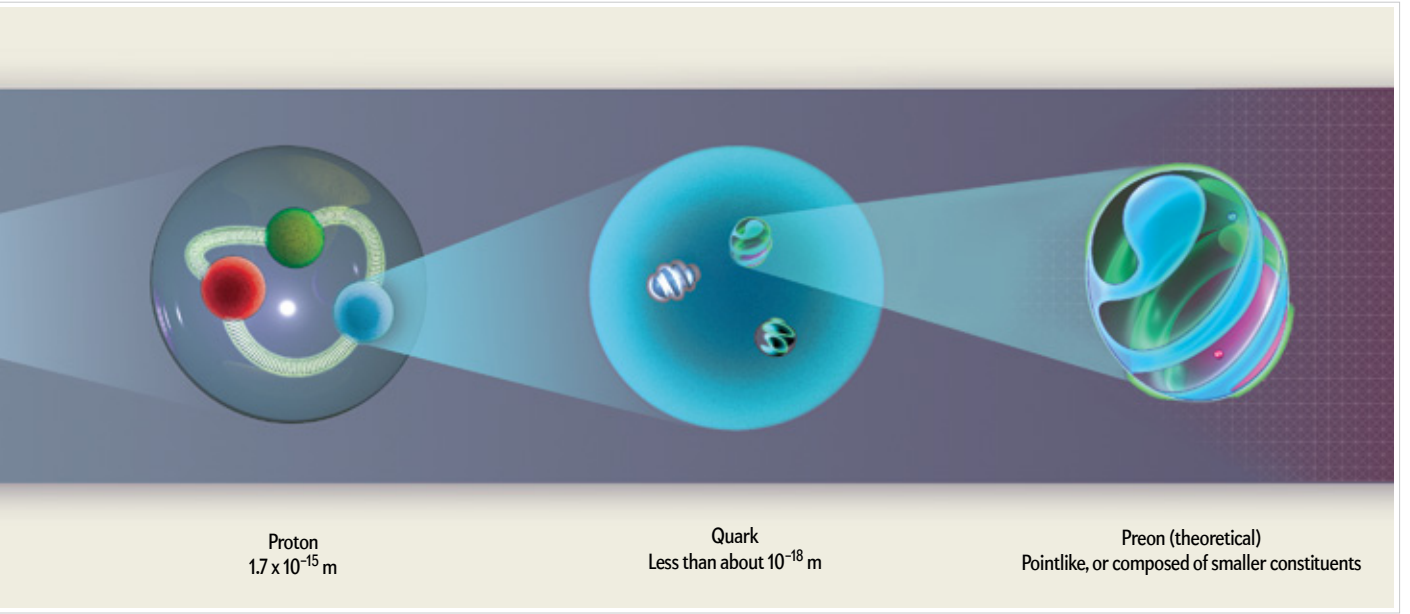
To distinguish between those two possibilities (zero size versus very, very tiny), we need to make more precise measurements. The LHC is a discovery machine, and the huge amount of data expected from its current collisions and a scheduled upgrade in the accelerator's energy are two ways in which we can expect to learn more about the size of quarks and leptons.

Another way to demonstrate the existence of particle substructure—for leptons, at least—is to investigate the tightly related concepts of spin and magnetic moment. With some poetic license, an electron can be thought of as a spinning ball, and physicists quantify that property with a spin quantum number. Like all fermions, the electron is said to have spin 1/2. Because the electron is electrically charged, the combination of spin and charge confers a magnetic moment, which is just a fancy way of saying it turns the electron into a familiar magnet, with a north and south pole. Assuming a lepton is a pointlike particle with spin 1/2, it should have a single and specific magnetic moment. So if a measurement of the electron or muon turns up a magnetic moment that differs from the prediction, that result would strongly suggest that those particles are not pointlike and therefore could be composed of preons.

Physicists have long known that the magnetic moments of both the electron and muon do diverge somewhat from that of a pointlike particle. This small difference has nothing to do with preons, though, and can actually be explained within the Standard Model. Each lepton is surrounded by an evanescent cloud of so-called virtual particles, which flicker in and out of existence. Because this cloud has a size, it alters the magnetic moment of the lepton ever so slightly—by about one part in 1,000. The effects of preons would be even smaller, but they could be detectable. New measurements on the horizon, at Fermilab's muon g-2 experiment, will be more than a factor of four more precise than those thus far achieved.

Physicists have also dug through collider data to look for particle decays that would be expected if preons exist and if the higher generations of particles are simply excited states of the first. One such process is a muon decaying into an electron and a photon. This decay has not yet been observed and, if it happens

© 2013 Scientific American



Proton
1.7×10^{-15} m

Quark
Less than about 10^{-18} m

Preon (theoretical)
Pointlike, or composed of smaller constituents

at all, occurs less than about one time in 100 billion.

Every direct measurement made to date is consistent with the hypothesis that quarks and leptons are, indeed, pointlike, with spin 1/2. For those of us who think that the observed generations of subatomic particles are a tantalizing hint of undiscovered physics, the past few decades have been frustrating. But now we have real opportunities to explore new territory.

In 2011 the LHC collided beams of protons at an energy of seven trillion electron volts (7 TeV), 3.5 times the previous world record (held by Fermilab's Tevatron for more than a quarter of a century). In that one year the LHC delivered half as much data from collisions as the Tevatron did over its entire 28-year operating career. In 2012 CERN raised the LHC's energy to 8 TeV before initiating a temporary shutdown at the end of the year to make repairs and improvements. The LHC is scheduled to resume operations in 2014, with a full program of colliding proton beams at 13 or 14 TeV beginning in early 2015.

The modest 2012 increase in energy meant a lot for preon searches. Improvements in accelerator operations increased the brightness and duration of the beams, quadrupling the number of collisions. In addition, the small increase in beam energy quintupled the number of highly energetic collisions—which probe the smallest sizes and which are exactly the kinds of events we need to inspect for evidence of preons. Thanks to these two improvements, the data from 2012 provided 20 times as many collisions of interest to preon hunters as in 2011 and have cut the maximum allowed size of preons in half. Ongoing analysis should enable us to again cut the maximum size in half—and perhaps even to find preons themselves! The next five years will be an exciting time for those of us looking for the ultimate building blocks of the universe.

In addition to the LHC, the Fermilab research program is undergoing a fundamental retooling, which will include a new ability to search for direct evidence of preons. Since the Tevatron was decommissioned in 2011, Fermilab's accelerators no longer tread the energy frontier of particle physics. Instead Fermilab is pushing forward into the intensity frontier, exploring rare phenomena with unprecedented precision. Two of the experiments are most relevant to the preon search. One will measure the magnetic moment of the muon, and the other will look for a muon that decays into a single electron without producing any neutrinos.

The future of hunting for structure within the quarks and leptons is brighter than it has been for a long time. As you read this article, my colleagues and I are combing through the huge amount of LHC data already taken. We are searching for evidence that quarks and leptons have a nonzero size. We are looking for a fourth generation of quarks and leptons and for some evidence that the force-carrying particles also have generations—that the W and Z bosons, which mediate the weak nuclear force, have heavier cousins.

The next few years will mark the start of a new foray into the subatomic realm, a journey the likes of which scientists last encountered more than 25 years ago, when the Tevatron began operations. Like intrepid adventurers of yesteryear, physicists are forging ahead and blazing a trail into the quantum frontier. SA

Don Lincoln, who has been interested in quark and lepton substructure for decades, is a senior physicist at Fermilab. He splits his research time between Fermilab and CERN and, in the occasional spare moment, also writes books and articles on particle physics for the public. He lives in the Chicago suburbs with his family and a particularly hirsute cat.

MORE TO EXPLORE

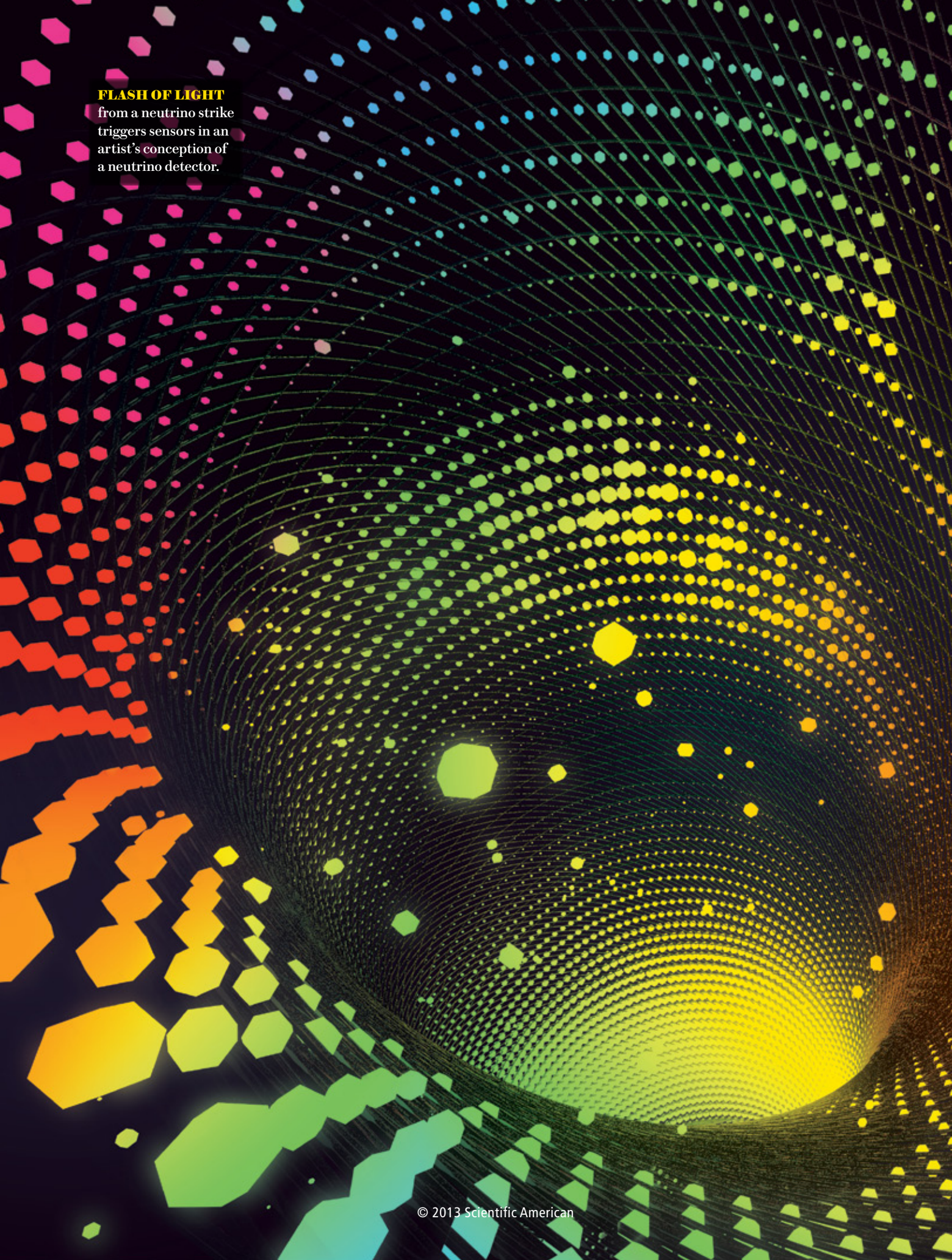
A Composite Model of Leptons and Quarks. Michael A. Shupe in *Physics Letters B,* Vol. 86, No. 1, pages 87–92; September 10, 1979.

A Schematic Model of Quarks and Leptons. Haim Harari in *Physics Letters B,* Vol. 86, No. 1, pages 83–86; September 10, 1979.

Preons: Models of Leptons, Quarks and Gauge Bosons as Composite Objects. Ian A. D'Souza and Calvin S. Kalman. World Scientific Publishing, 1992.

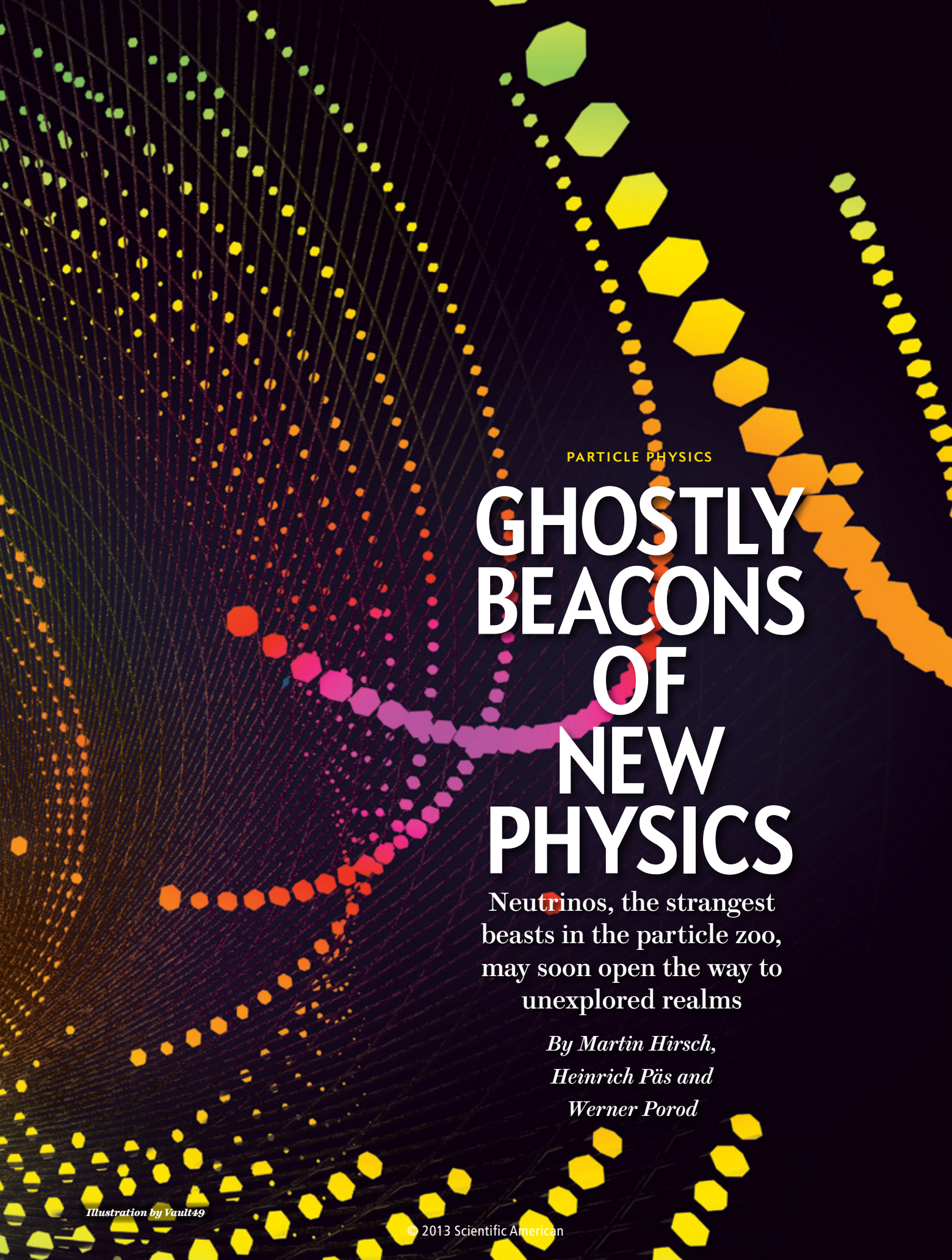
The Quantum Frontier: The Large Hadron Collider. Don Lincoln. Johns Hopkins University Press, 2009.

© 2013 Scientific American



FLASH OF LIGHT from a neutrino strike triggers sensors in an artist's conception of a neutrino detector.

© 2013 Scientific American



PARTICLE PHYSICS

GHOSTLY BEACONS OF NEW PHYSICS

Neutrinos, the strangest beasts in the particle zoo, may soon open the way to unexplored realms

By Martin Hirsch, Heinrich Päs and Werner Porod

Illustration by Vault49

© 2013 Scientific American

FEW PHYSICISTS HAVE HAD THE PRIVILEGE OF BRINGING A NEW ELEMENTARY PARTICLE into the world. When Wolfgang Pauli hit on the idea of the neutrino in 1930, however, internal misgivings tempered his response. "I have done a terrible thing," Pauli later told his colleagues. "I have postulated a particle that cannot be detected." The neutrino is indeed elusive—its ghostly nature allows it to slip through almost all physical barriers, including the materials that physicists use in their particle detectors. In fact, most neutrinos pass cleanly through the earth without so much as brushing against another particle. Yet Pauli's fears turned out to be slightly overblown: the neutrino can be detected—although doing so requires great effort and experimental ingenuity.

Neutrinos are the oddest of the fundamental particles on other grounds as well. They do not make up atoms, nor do they have anything to do with chemistry. They are the only electrically neutral matter particles. They are extremely light—less than a millionth the mass of the next-to-lightest matter constituent, the electron. And neutrinos, more than other particles, metamorphose; they shift among three varieties, or "flavors."

These tiny particles have kept physicists in continuous astonishment for more than 80 years. Even today fundamental questions about the neutrino remain unanswered: Are there only three flavors of neutrino, or do more exist? Why are all neutrinos so lightweight? Are neutrinos their own antimatter counterparts? Why do neutrinos shift character with such amazing verve?

Around the world—at particle colliders, at nuclear reactors, in abandoned mine shafts—new experiments that can address these questions are coming online. The answers they deliver should provide essential clues to the inner workings of nature.

The neutrino's oddities make it a lodestar guiding particle physicists on the daunting voyage toward a so-called grand unified theory describing all particles and forces, except gravity, in a consistent mathematical framework. The Standard Model of particle physics, the best theory of particles and forces to date, cannot accommodate all the complexities of the neutrino. It must be extended.

LIGHTWEIGHT BUT PRESSING

THE MOST POPULAR WAY to build on the neutrino segment of the Standard Model is to introduce new entities called right-handed neutrinos. Handedness is a variant of electrical charge that determines whether a particle feels the weak interaction, the force responsible for radioactive decay; a particle must be left-handed to feel the weak force. These hypothetical right-handed particles would thus be even slipperier than their left-handed fel-

IN BRIEF

The neutrino is the oddest breed of fundamental particle. Neutrinos seem to defy all precedents set by better understood varieties of particles, such as electrons and quarks.

Lightweight, shifty and exceedingly difficult to detect, neutrinos have been vexing experimentalists for decades.

Even today fundamental properties of neutrinos remain up for debate. Some of the key questions pertain to the origin of their meager masses, the nature of neutrino antimatter and the number of neutrino species in existence, not to mention their penchant for switching identities on the fly.

Uncovering the true nature of the neutrino may pave the way to a more unified theory of physics.

© 2013 Scientific American

lows, the experimentally proved neutrinos of the Standard Model. All neutrinos are classified as leptons—the extended family of particles that also includes the electrons—meaning that they do not feel the strong force holding together the protons and neutrons in the atomic nucleus. Lacking electrical charge, neutrinos do not directly feel electromagnetic forces, either. That leaves only the force of gravity and the weak interaction for the three known neutrino flavors, but a right-handed neutrino would be impervious even to the weak force.

If a right-handed neutrino exists, it would provide a very reasonable explanation for another neutrino puzzle: the reason the three left-handed varieties—the electron, muon and tau neutrinos—all have such tiny masses.

Most elementary particles gain their mass by interacting with the ubiquitous Higgs field. (Higgs made headlines in March when physicists at the Large Hadron Collider, or LHC, at CERN near Geneva announced results strongly indicating that a new particle identified in 2012 was indeed the long-sought Higgs boson. That boson is the particle counterpart to the Higgs field, just as the photon is the counterpart to the electromagnetic field.) In the process, the Higgs carries away the particles' weak force version of electrical charge. Because right-handed neutrinos lack this charge, their mass does not rely on the Higgs field. Instead it may emerge from a different mechanism altogether at the extremely high energies of grand unification, which would make the right-handed neutrino enormously heavy.

Quantum effects could link right-handed neutrinos to their left-handed siblings in a way that would cause the enormous mass of one to "infect" the other. The contagiousness would be very weak, though—if the right-handed neutrino came down with pneumonia, the

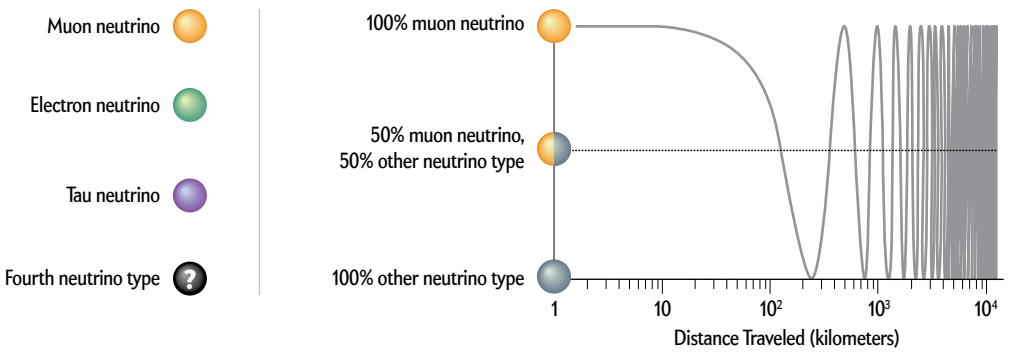
BASICS

Changing Identities on the Fly

As neutrinos propagate, at nearly the speed of light, through space, the earth or your body, they change identities often, oscillating between three known neutrino types. Their behavior is odd but not entirely random—the properties of neutrinos allow physicists to predict the probability of their oscillation over various distances.

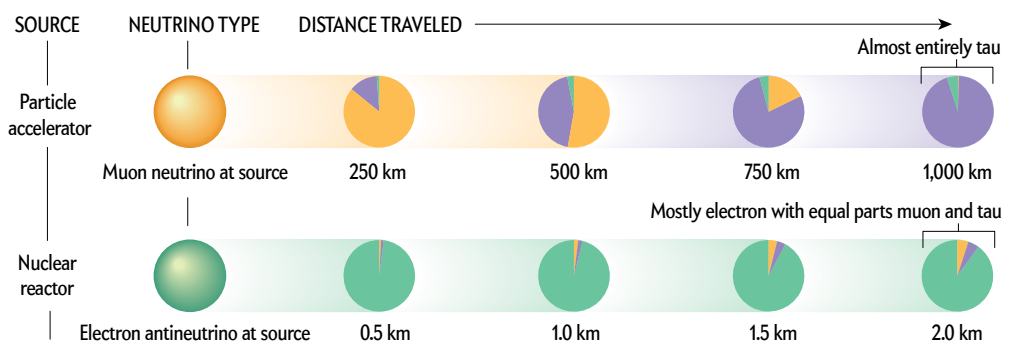
Flavor Palette

Neutrinos come in at least three flavors: the electron, the muon and the tau. (Some physicists suspect the existence of a fourth neutrino as well.) The graph below illustrates the probability that a muon neutrino, having traversed a given distance, will switch flavors. In practice, oscillation distances depend on the neutrino's energy.



Classic Morphs

Neutrino oscillation experiments measure the flavor divergence between the particles emitted by a neutrino source and those captured at a distant detector. The illustration below shows idealized source-to-detector oscillation patterns for experiments at particle accelerators and nuclear reactors.



Graphic by Jen Christiansen

© 2013 Scientific American

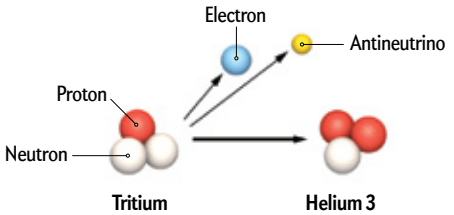
EVIDENCE

The Heart of the Antimatter

A number of experiments around the globe have been built to observe a rare nuclear physics phenomenon called double beta decay. These experiments aim to test the hypothesis that, within the realm of neutrinos, matter and antimatter are one and the same. If neutrinos are indeed their own antiparticles, they could alter the balance of matter and antimatter, potentially explaining how matter came to dominate the universe.

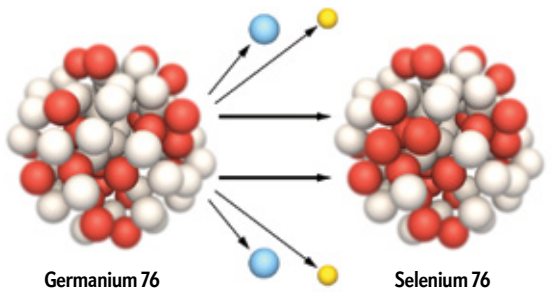
Single Beta Decay

Radioactive nuclei can settle into more stable configurations via beta decay. Here tritium (an isotope of hydrogen) decays to helium 3 by transmuting a neutron into a proton, shedding an electron and an antineutrino in the process.



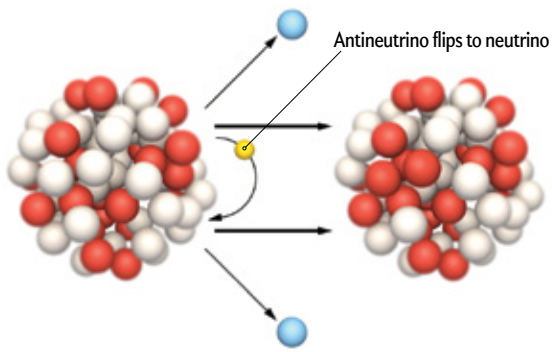
Double Beta Decay

A handful of nuclear isotopes can undergo two beta decays at once—switching two neutrons into protons, which emits two electrons and two antineutrinos.



Neutrinoless Double Beta Decay

If neutrinos are their own antiparticles, the neutrino part of the decays could cancel out—in one decay a neutrino would be absorbed instead of an antineutrino being emitted. Such neutrinoless decay has not yet been convincingly observed.



left-handed one would catch only a minor cough—which means the left-handed mass would be very tiny. This relation is known as the seesaw mechanism because a large mass raises, or lifts up, a smaller mass.

An alternative explanation for neutrino masses arises from supersymmetry, a leading candidate for new physics beyond the Standard Model. In the supersymmetry hypothesis, every particle within the Standard Model has an as yet undiscovered partner. The so-called superpartner particles, which must be exceedingly massive to have thus far escaped detection, would instantly at least double the number of elementary particles. If supersymmetric particles exist, the LHC may be able to produce them and measure their properties.

One of the most appealing traits of supersymmetry is that a superparticle known as the neutralino makes a fine candidate for dark matter—the mass in galaxies and galaxy clusters that exerts a gravitational pull but does not emit light nor reveal itself in other obvious ways. The neutralino would fit the bill for dark matter only if it were stable over long periods, rather than decaying away rapidly to some other particle.

A short-lived neutralino would thus send dark matter researchers back to the drawing board but could prove a boon to neutrino physicists. The neutralino's stability depends on a hypothetical property called R-parity, which prevents the superpartners from decaying into any of the ordinary Standard Model particles. If R-parity does not hold, however, the neutralino becomes unstable—and its decay depends in part on the mass of the neutrino.

Two of us (Hirsch and Porod), in collaboration with José Valle of the University of Valencia in Spain and Jorge C. Romão of the Technical University of Lisbon in Portugal, have shown that the link between neutrinos and the neutralino could be testable at the LHC. If the stability of the neutralino indeed depends on neutrinos, the neutralino's lifetime would be predictable from known neutrino properties. And it just so happens that the superparticle should exist long enough for physicists to track its entire lifetime—from production to decay—inside the detectors of the LHC.

WHAT'S THE ANTIMATTER?

ALL PLAUSIBLE EXPLANATIONS for the neutrino's meager masses point to unexplored realms of physics. Yet one of those explanations, the seesaw mechanism, may also bear on the mystery of how matter came to reign over antimatter—a triumph that enabled the formation of cosmic structure and, ultimately, the development of life.

Every particle in the Standard Model has an antimatter counterpart, a sort of Bizarro World version with an opposing charge. The electron, for instance, has an electrical charge of −1, and the antielectron, or positron, has a charge of +1. When an electron and positron collide, their charges cancel out, and the particles annihilate in a burst of radiation. The complete chargelessness of the right-handed neutrino may have an important consequence: it could mean that, for neutrinos, matter and antimatter are one and the same. In the terminology of physics, the electron and positron are known as Dirac particles. A particle that is its own antimatter counterpart, on the other hand, is a Majorana particle.

If the seesaw theory accurately reflects the workings of

Illustrations by George Retseck

© 2013 Scientific American

the particle world, then the left-handed neutrinos are infected not just with mass but also with the Majorana-ness of the right-handed neutrinos. In other words, if some neutrinos are their own antiparticles, then all neutrinos are.

Neutrinos and their antiparticles being one and the same would have a variety of fascinating implications. For instance, neutrinos could trigger transitions among particles and antiparticles. In most particle reactions, the so-called lepton number, or the number of leptons minus the number of antileptons, is conserved—it does not change. Neutrinos, however, might violate this rule, creating an imbalance of matter and antimatter. For us humans, the imbalance is a very good thing because if matter and antimatter were equally paired in the aftermath of the big bang, they would have completely annihilated each other and left nothing behind to build galaxies, planets and life-forms. The explanation for matter's dominance over antimatter has long eluded physicists and cosmologists.

DISAPPEARING ACT

THE CONNECTION between neutrinos and their antiparticles does not have to languish in the realm of tantalizing but ultimately unsettled theory. Many experiments, past and present, have sought to answer definitively whether neutrinos are in fact their own antiparticles by searching for a type of radioactive event known as nuclear double beta decay.

Neutrinos and antineutrinos were first observed in nuclear beta decay, by which an atom emits an electron, along with an antineutrino. In several nuclear isotopes, two beta decays can occur simultaneously, which, under normal circumstances, emits two electrons and two antineutrinos. Yet if the neutrino is a Majorana particle, then the same antineutrino emitted in the first decay can be absorbed in the second. The result is a double beta decay that does not emit any neutrinos or antineutrinos [*see box on opposite page*]. In an instant, where there had previously been no leptons, two leptons (the electrons) emerge without their usual, counterbalancing antileptons (the antineutrinos). In other words, this so-called neutrinoless double beta decay violates the conservation of the lepton number.

At present, the search for neutrinoless double beta decay is the best test we have for Majorana neutrinos in particular and for lepton number violation in general. In principle, a neutrinoless double beta decay experiment is simple: collect a nuclear isotope such as germanium 76, in which simultaneous beta decays can occur, and wait for the emergence of two electrons unaccompanied by neutrinos. In practice, however, the experiments are very difficult. Double beta decay of any kind is exceedingly rare, so experimenters must gather large quantities of germanium, or other source materials, to have a hope of documenting the neutrinoless variety. To make matters worse, the constant stream of subatomic particles raining down on the earth from cosmic rays tends to drown out the minuscule signal from double beta decays. So experimentalists must bury their detectors deep underground or in former mines or other subterranean labs, where the overlying rock screens out nearly all cosmic rays.

At particle colliders, at nuclear reactors and in abandoned mine shafts, next-generation neutrino experiments are now coming online. They should provide essential clues to the inner workings of nature.

Unfortunately, the only report to date of neutrinoless double beta decay, from the Heidelberg-Moscow Double Beta Decay Experiment in Italy, has been vigorously contested by other physicists. Next-generation detectors just starting to take data or currently under construction will conduct a more thorough search. An experiment in New Mexico called EXO-200 and one in Japan named KamLAND-Zen recently published the first data from their searches for neutrinoless double beta decay, which caused friction with the earlier claim but did not unambiguously rule it out.

The GERDA experiment in Italy, which came online in 2011, uses the same isotope as the Heidelberg-Moscow setup in an improved design that aims to directly confront its predecessor's controversial finding. Both the EXO-200 and KamLAND-Zen experiments are continuing their operations, and an apparatus known as CUORE is scheduled to start taking data in Italy in 2014. The number of advanced experiments now under way provides a very reasonable hope that neutrinoless double beta decay may be confirmed before the end of this decade.

LIGHT SWITCHERS

FINDING AN AS YET UNDISCOVERED neutrino or proving that neutrinos and antineutrinos are one and the same would add an entirely new layer of intrigue to these already puzzling particles. But even as we physicists hunt for new facets of these particles, we continue to wrestle with the mechanism underlying a well-documented but poorly understood attribute of neutrinos—their strong propensity to metamorphose. In the literature, we say that the amount of lepton flavor violation, or neutrino mixing, is large in comparison with the mixing among flavors of quarks, the elementary particles that make up protons and neutrons.

Many research groups worldwide are investigating how newly conceived symmetries of nature—key commonalities between apparently distinct forces and particles—could explain such behavior. One example would be the symmetries inherent in the ways that the known particles transform from one to another. Gautam Bhattacharyya of the Saha Institute of Nuclear Physics in Calcutta, Philipp Leser of the Technical University of Dortmund in Germany and one of us (Päs) recently discovered that such symmetries would conspicuously affect the Higgs field. The interaction of flavor-swapping quarks and neutrinos with the Higgs field would manifest itself in exotic decay products of Higgs bosons that ought to be observable at the LHC. Such a signal could point to the underlying mechanism for neutrinos' hyperactive transmutations, which would certainly be one of the most spectacular discoveries at the LHC.

In the meantime, a different family of experiments is nailing down just how often the particles switch identities. Long-baseline experiments such as T2K in Japan, MINOS in Minnesota and OPERA in Italy detect beams of neutrinos that

© 2013 Scientific American

originate at particle accelerators hundreds of kilometers away, to measure changes in flavor as neutrinos traverse long distances through the earth [*see box on page 23*]. The scales of these experiments are so large that the neutrinos may cross state lines or even international borders on their journeys. (In 2011 OPERA made news when physicists from the collaboration announced that neutrinos in their experiment appeared to travel from CERN to an underground Italian lab faster than the speed of light—a measurement that soon proved to be flawed.) In complement to these long-distance neutrino experiments, the Double Chooz project in France, the Daya Bay Reactor Neutrino Experiment in China and RENO in South Korea all measure the short-range oscillation of neutrinos coming from nuclear reactors.

Only in 2012 did these experiments finally determine the last and smallest of the so-called mixing angles—the pa-

MEASURING MASS

The Neutrino's Secrets, Written on the Sky

By Sudeep Das and Tristan L. Smith

Measuring the minuscule mass of neutrinos has so far proved impossible—and not for lack of trying. Numerous laboratory experiments over the past few decades have succeeded only in placing loose limits on the three neutrino masses.

We have very compelling reasons to expect that the best way to measure the mass of these tiny particles is, surprisingly, to look for their influence at the largest scales of the universe. For although neutrinos are virtually massless and nearly invisible, their sheer numbers—some 10^{89} in the universe—make them very consequential players in the cosmos.

Our logic goes like this: Early in the history of the universe, when everything was very hot and dense, nuclear reactions forged helium from hydrogen, releasing huge numbers of neutrinos as a by-product. As the universe evolved, expanded and cooled, small fluctuations in the density of this primordial particle soup were amplified; in regions of above-average density, gravity tried to pull more material in.

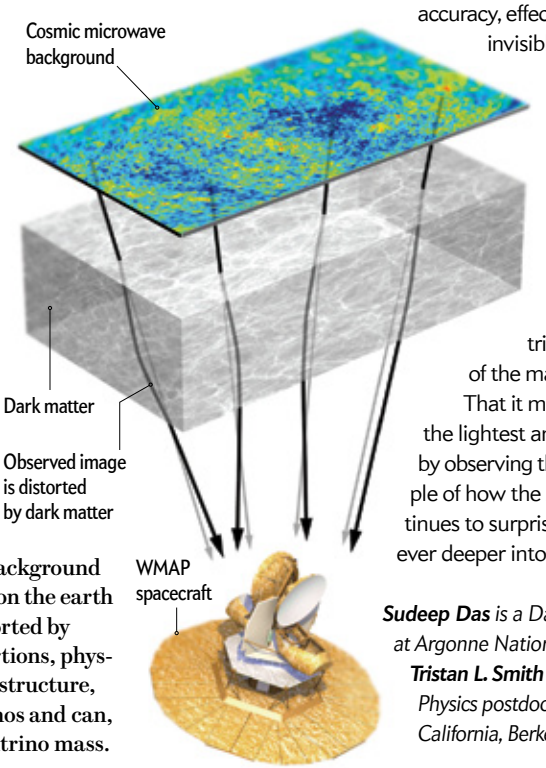
Dark matter, the essentially invisible stuff that accounts for a goodly portion of the universe's mass, collapsed into clumps first because it only interacts through gravity. These initial clumps of dark matter formed the seeds of the galaxies and the clusters of galaxies that we see today. Neutrinos, being extremely light, began to clump somewhat later on in the universe's development. In fact, by zipping so freely through the cosmos, neutrinos have actually slowed the clumping of dark matter—an effect that should be detectable today.

The more massive the neutrinos, the more they will have hindered the clumping of matter—in effect, blurring the edges in the large-scale structure of the universe. By measuring how matter is distributed in the universe, then, we can infer how massive neutrinos are.

Mapping the distribution of matter—most of which is dark matter—is hardly straightforward. Yet researchers have seen that the leftover radiation from the big bang, known as the cosmic microwave background (CMB), is slightly distorted because of the light-bending gravitational effects of dark matter clumps that fill the space between the CMB and us. Examining this "gravitational lensing" of the CMB is a very promising way of measuring the distribution of dark matter in the universe.

New precision measurements of the CMB now under way will enable us to measure the lensing distortions to very high accuracy, effectively mapping out the otherwise invisible dark matter. If the distribution of dark matter is confined to sharp-edged structures separated by voids, we can infer that the neutrino masses are small; if instead the edges are blurred, we will know that the neutrino masses are larger. The new generation of CMB experiments should allow us to pin down the combined masses of the three neutrino types to within one five-millionth of the mass of the electron.

That it may be possible to measure the mass of the lightest and most elusive of subatomic particles by observing the entire universe is just another example of how the study of physics, across all scales, continues to surprise and inspire astrophysicists to delve ever deeper into the workings of the natural world.



WARPED: Cosmic microwave background radiation collected by telescopes on the earth and in space has been subtly distorted by dark matter. By tracing the distortions, physicists can chart the dark matter's structure, which has been shaped by neutrinos and can, in turn, place strict limits on neutrino mass.

Sudeep Das *is a David Schramm postdoctoral fellow at Argonne National Laboratory.*

Tristan L. Smith *is a Berkeley Center for Cosmological Physics postdoctoral fellow at the University of California, Berkeley.*

© 2013 Scientific American

rameters governing the transitions among neutrino flavors. The final mixing angle to be pinned down, known as the reactor angle, describes the probability of an electron neutrino or antineutrino's conversion over a short baseline. The measurements of the reactor angle opened the possibility that future neutrino experiments might be able to compare the properties of neutrinos and antineutrinos. An asymmetry between particles and their antimatter counterparts would be known as CP violation and, along with studies of neutrinoless double beta decay, could bear on the mystery of why there is more matter than antimatter in our universe.

Of the ongoing searches, T2K probably has the first decent chance to see hints of CP violation. But the race is on among this new generation of experiments to answer key neutrino questions—and it promises to be exciting. The long-baseline NOvA experiment, now under construction in the U.S., also has the potential to uncover CP violation in neutrinos. NOvA will fire a neutrino beam through the earth from Fermi National Accelerator Laboratory in Batavia, Ill., clear across the state of Wisconsin and the tip of Lake Superior, to a detector in Ash River, Minn., 810 kilometers away. The neutrinos will make the trip in less than three milliseconds.

Among its research goals, NOvA also aims to clarify the neutrino mass hierarchy—determining which of the neutrinos is the lightest and which is the heaviest. At present, physicists know only that at least two neutrino species have nonzero masses, but, as with so many aspects of these ghostly particles, the details elude us.

LINGERING MYSTERIES

WITH SO MANY NEUTRINO EXPERIMENTS under way—featuring different aims, different designs and different particle sources—the varied data emerging from around the globe have sometimes yielded conflicting interpretations. One of the most tantalizing—as well as controversial—experimental hints suggests the existence of a new particle called the sterile neutrino.

Echoing Pauli's fears in 1930, the sterile neutrino would be only indirectly detectable, just like the much heavier right-handed neutrino of the seesaw mechanism. (From a theoretical point of view, however, the two proposed particles are nearly mutually exclusive.) Nevertheless, two experiments may have caught a whiff of the sterile neutrino. LSND, which ran at Los Alamos National Laboratory in the 1990s, found early but controversial evidence for an elusive type of neutrino flavor conversion—muon antineutrinos morphing into electron antineutrinos. Fermilab's MiniBooNE, which began producing scientific results in 2007, also hinted at such conversions. Yet the LSND and MiniBooNE oscillations did not fit neatly into the standard three-neutrino picture.

Any asymmetry between neutrinos and their antimatter counterparts could bear on a mystery that has long vexed physicists and cosmologists: why there is more matter than antimatter in our universe.

Quantum mechanics permits neutrinos to oscillate between flavors only if they have mass—and only if each flavor has a different mass. The various neutrino masses could trigger neutrino conversion to explain the LSND and MiniBooNE anomalies but only if another mass difference exists in addition to the ones already known—in other words, only if four neutrino types exist instead of three. An additional neutrino coupling to the weak force would make the Z boson—a carrier of the weak force—decay too fast, so this particle would not interact with the weak force at all. Hence the "sterile" designation: this hypothetical neutrino would be almost entirely decoupled from the rest of the particle zoo.

Detectors of a different kind altogether, which capture neutrinos from nearby nuclear reactors, have also registered surprising results that could point to a sterile neutrino. The data from several reactor experiments indicate an anomalous disappearance of electron antineutrinos over very short distances, which, if interpreted in terms of neutrino oscillations, would imply the existence of sterile neutrinos. The anomaly has been around for some time, but recent recalculations of the neutrino output from the various reactors have strengthened the case for a new particle.

The evidence for sterile neutrinos, such as it is, remains sketchy, indirect and conflicted—all of which is to be expected in pursuit of a notoriously elusive, and possibly nonexistent, particle. Yet MiniBooNE and a companion experiment called MicroBooNE, which is now under construction at Fermilab, may soon have something firmer to say on the matter. And a new crop of proposed experiments, which would study the reactor anomaly, is also under discussion.

It is remarkable that the mighty LHC and the comparatively low-energy experiments on the humble neutrino provide such complementary routes to explore the inner workings of nature. More than 80 years after Wolfgang Pauli conceived of his "particle that cannot be detected," neutrinos continue to guard their secrets closely. Still, the potential payoff in unraveling those secrets justifies the decades-long effort to pry ever further into the neutrino's private life. SA

Martin Hirsch is a professor in the Astroparticles and High-Energy Physics Group at IFIC, a joint center for particle physics at the University of Valencia and the Spanish National Research Council.
Heinrich Päs is a professor at the Technical University of Dortmund in Germany. *The Perfect Wave*, his book on neutrinos, is forthcoming from Harvard University Press.
Werner Porod is a professor at the University of Würzburg in Germany.

MORE TO EXPLORE

Testing Neutrino Mixing at Future Collider Experiments. W. Porod, M. Hirsch, J. Romão and J.W.F. Valle in *Physical Review D*, Vol. 63, No. 11, Article No. 115004; April 30, 2001.
Physics of Neutrinos: And Applications to Astrophysics. Masataka Fukugita and Tsutomu Yanagida. Springer, 2003.
Probing Minimal Supergravity in Type-I Seesaw Mechanism with Lepton Flavor Violation at the CERN LHC. M. Hirsch, J.W.F. Valle, W. Porod, J. C. Romão and A. Villanova del Moral in *Physical Review D*, Vol. 78, No. 1, Article No. 013006; July 18, 2008.
Exotic Higgs Boson Decay Modes as a Harbinger of S_3 Flavor Symmetry. Gautam Bhattacharyya, Philipp Leser and Heinrich Päs in *Physical Review D*, Vol. 83, No. 1, Article No. 011701; January 7, 2011.
Neutrinoless Double-Beta Decay and Physics beyond the Standard Model. Frank F. Deppisch, Martin Hirsch and Heinrich Päs in *Journal of Physics G: Nuclear and Particle Physics*, Vol. 39, No. 12, Article No. 124007; November 19, 2012.

© 2013 Scientific American

QUANTUM PHYSICS

LOOPS, TREES AND THE SEARCH FOR NEW PHYSICS

EXTREME PHYSICS

Maybe unifying the forces of nature isn't quite as hard as physicists thought it would be

By Zvi Bern, Lance J. Dixon and David A. Kosower

Illustration by Kenn Brown and Chris Wren, Mondolithic Studios

© 2013 Scientific American

© 2013 Scientific American

EXTREME PHYSICS

ON A SUNNY SPRING DAY ONE OF US (DIXON) ENTERED THE London Underground at the Mile End station on his way to Heathrow Airport. Eyeing a stranger, one of more than three million daily passengers on the Tube, he idly wondered: What is the probability the stranger would emerge at, say, Wimbledon? How could you ever figure that out, given that the person could take any number of routes? As he thought about it, he realized that the question was similar to the knotty problems that face particle physicists who seek to make predictions for particle collisions in modern experiments.

The Large Hadron Collider (LHC) at CERN near Geneva, the premier discovery machine of our age, smashes together protons traveling at nearly the speed of light to study the debris from their collisions. Building the collider and its detectors pushed technology to its limits. Interpreting what the detectors see is an equally great, if less visible, challenge. At first glance, that seems rather strange. The Standard Model of elementary particles is well established, and theorists routinely apply it to predict the outcomes of experiments. To do so, we rely on a calculational technique developed more than 60 years ago by the renowned physicist Richard Feynman. Every particle physicist learns Feynman's technique in graduate school. Every book and magazine article about particle physics for the public is based on Feynman's concepts.

Yet his technique has become obsolete for state-of-the-art problems. It provides an intuitive, approximate way to grasp the very simplest processes but is hopelessly laborious for more complicated ones or for high-precision calculations. Predicting what will emerge from a particle collision is even more daunting than predicting where a subway passenger will go. All the computers in the world working together would be unable to determine the outcome of even a fairly common collision at the LHC. If theorists cannot make precise predictions for known laws of physics and known forms of matter, what hope do we have of telling when the collider has seen something truly new? For all we know, the LHC may already have found answers to some of the greatest mysteries of nature, and we remain in the dark just because we cannot solve the equations of the Standard Model accurately enough.

In recent years the three of us and our colleagues have developed a new way of analyzing particle processes that bypasses the complexity of Feynman's technique. Called the unitarity method, it amounts to a highly economical way of predicting what a subway passenger will do by realizing that the passenger's options at each decision point are actually rather limited and can be broken down into probabilities for sequences of actions. Many theoretical problems in particle physics that were intractable have been cracked wide open by the new idea. Their solutions allow us to understand in unprecedented detail what our current theory of nature predicts so that we will know a new discovery when we see it. The method has also produced a wealth of results for an idealized cousin of the Standard Model that is of special interest to physicists as a stepping-stone to the ultimate theory of nature.

The unitarity approach is more than a helpful calculational trick. It suggests a radical new vision of theories of particle interactions that are governed by unexpected symmetries, reflecting an underappreciated elegance of the Standard Model. Notably, it has revealed a strange

IN BRIEF

Recently a silent revolution has swept through physicists' understanding of particle collisions. The concepts introduced by the iconic physicist Richard Feynman have reached the limit of their usefulness for many applications, and the authors and their colleagues have developed a fresh approach.
Using it, physicists can describe more reliably how ordinary particles behave under the extreme conditions at the Large Hadron Collider at CERN, aiding experimentalists in their search for exotic particles and forces.
More profoundly, the novel methods breathe new life into a unified theory that physicists left for dead in the 1980s. The force of gravity looks like two copies of the strong subnuclear interactions working in unison.

© 2013 Scientific American

twist in the decades-old effort to unite quantum theory and Einstein's general theory of relativity into a quantum theory of gravity. Up until the 1970s, physicists assumed that gravity behaves like the other forces of nature and sought to extend our existing theories to cover it. When they applied Feynman's technique, however, they either got nonsensical results or were stymied by the math. Gravity, it seemed, was not like the other forces after all. Discouraged, physicists turned to more revolutionary ideas such as supersymmetry and, later, string theory.

The unitarity method, however, has allowed us to actually do calculations that were contemplated in the 1980s but seemed hopelessly beyond reach then. We have found that some of the supposed inconsistencies are in fact absent. Gravity does look like the other forces, albeit in an unexpected way—it behaves like a "double copy" of the strong subnuclear force that binds the constituents of nuclei together. The strong force is transmitted by particles known as gluons; gravity should be transmitted by particles known as gravitons. The new picture is that each graviton behaves like two gluons stitched together. This concept is quite strange, and even experts do not yet have a good mental image of what it means. Nevertheless, the double-copy property provides a fresh perspective on how gravity might be unified with the other known forces.

FROM TREES TO THICKETS

WHAT MADE FEYNMAN'S technique so compelling and useful is that it gives a precise graphical recipe for extremely complicated calculations. It is based on diagrams that give a visual picture of two or more particles colliding or scattering off one another. At every research institution studying elementary particle physics, you can find blackboards covered by these diagrams. To make quantitative predictions, a theorist draws a set of diagrams, each representing one of the conceivable ways a collision may unfold; it is analogous to one of the possible routes an Underground rider might take. Following a set of detailed instructions that Feynman and his colleagues, notably Freeman Dyson, laid down, the theorist then assigns a number to each diagram, giving the probability the event will take place in that way.

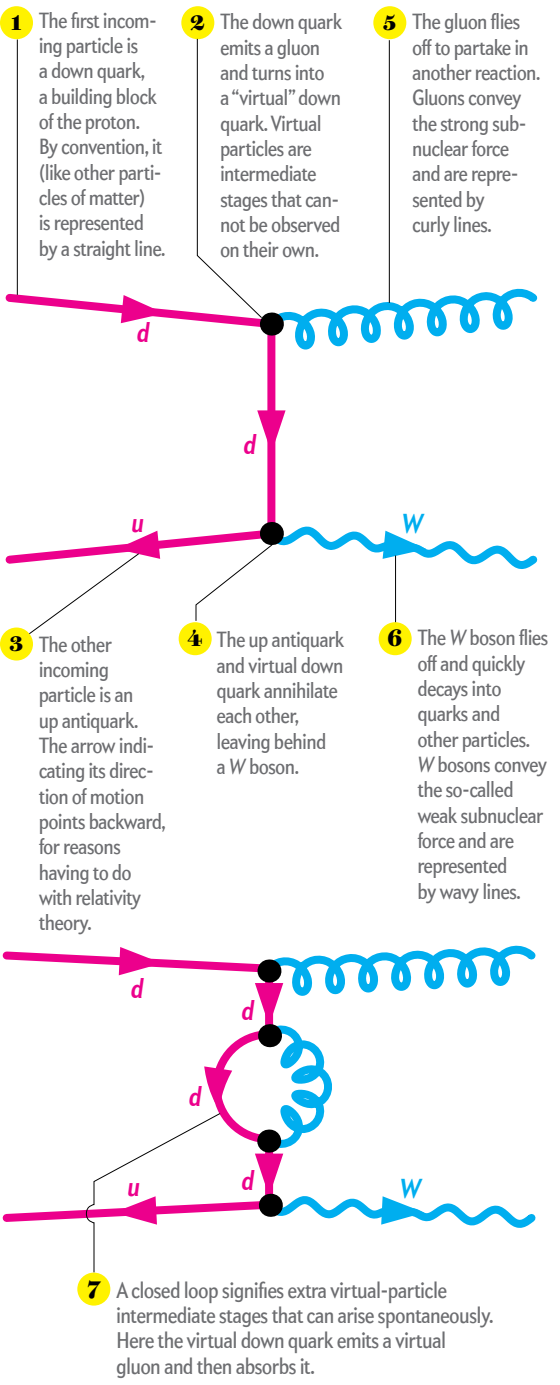
The downside is that the number of diagrams one can draw is enormous—in principle, infinite. In the applications for which Feynman originally developed his rules, this disadvantage did not matter. He was studying quantum electrodynamics (QED), which describes how electrons interact with photons. The interaction is governed by a quantity, the coupling, approximately equal to $1/137$. The smallness of the coupling ensures that complicated diagrams receive a low weighting in the calculation and can often be ignored altogether. That is like saying that an Underground rider is usually better off taking a fairly simple route.

Two decades later physicists extended Feynman's technique to the strong subnuclear force. By analogy with QED, the theory of the strong force is known as quantum chromodynamics (QCD). QCD is also governed by a coupling, but as the word "strong" suggests, its value is higher than that of the electromagnetic coupling. On the face of it, a larger coupling increases the number of complicated diagrams that theorists must consider—like an Underground rider who is willing to take very circuitous routes, making it hard to predict what he or she will do. Fortunately, at very short distances, including the distances relevant for collisions at the LHC, the coupling diminishes in val-

ANATOMY OF A FEYNMAN DIAGRAM

When Particles Collide

All around us all the time, elementary particles are attracting, repelling, hitting, mutating or annihilating one another. To visualize the quantum hurly-burly, Nobelist Richard Feynman created a system of stick-figure-like diagrams. Those below depict two quarks interacting to give birth to a gluon and *W* boson.


1 The first incoming particle is a down quark, a building block of the proton. By convention, it (like other particles of matter) is represented by a straight line.
2 The down quark emits a gluon and turns into a "virtual" down quark. Virtual particles are intermediate stages that cannot be observed on their own.
5 The gluon flies off to partake in another reaction. Gluons convey the strong subnuclear force and are represented by curly lines.
d
d
u
W
3 The other incoming particle is an up antiquark. The arrow indicating its direction of motion points backward, for reasons having to do with relativity theory.
4 The up antiquark and virtual down quark annihilate each other, leaving behind a W boson.
6 The W boson flies off and quickly decays into quarks and other particles. W bosons convey the so-called weak subnuclear force and are represented by wavy lines.
d
d
d
u
d
W
7 A closed loop signifies extra virtual-particle intermediate stages that can arise spontaneously. Here the virtual down quark emits a virtual gluon and then absorbs it.

Graphics by Zvi Bern, Lance J. Dixon, David A. Kosower and Harald Ita

© 2013 Scientific American

ue and, for the very simplest collisions, theorists can again get away with considering only uncomplicated diagrams.

For messy collisions, though, the full complexity of the Feynman technique comes rushing in. Feynman diagrams are classified by the number of external lines and closed loops they have [*see box on opposite page*]. Loops represent one of the quintessential features of quantum theory: virtual particles. Though not directly observable, virtual particles have a measurable effect on the strength of forces. They obey all the usual laws of nature, such as the conservation of energy and of momentum, with one caveat: their mass can differ from that of the corresponding "real" (that is, directly observed) particles. Loops represent their ephemeral life cycle: they pop into existence, move a short distance, then vanish again. Their mass determines their life expectancy: the heavier they are, the shorter they live.

The simplest Feynman diagrams ignore virtual particles; they have no closed loops and are called tree diagrams. In quantum electrodynamics, the simplest diagram of all shows two electrons repelling each other by exchanging a photon. Progressively more complicated diagrams add loops one by one. Physicists refer to this additive procedure as "perturbative," meaning that we start with some approximate estimate (represented by the tree diagrams) and gradually perturb it by adding refinements (the loops). For instance, as the photon travels between the two electrons, it can spontaneously split into a virtual electron and virtual antielectron, which live a short while before annihilating each other, producing a photon. The photon resumes the journey the original photon had been taking. In the next level of complexity, the electron and antielectron might themselves split temporarily. With increasing numbers of virtual particles, the diagrams describe quantum effects with increasing precision.

Even tree diagrams can be challenging. In the case of QCD, if you were brave enough to consider a collision involving two incoming and eight outgoing gluons, you would need to write down 10 million tree diagrams and calculate a probability for each. An approach called recursion, pioneered in the 1980s by Frits Berends of Leiden University in the Netherlands and Walter Giele, now at Fermilab, tamed the problem for tree diagrams but had no obvious extension to loops. Worse, closed loops make the workload overwhelming. Even a single loop causes an explosion in both the number of diagrams and the complexity of each. The mathematical formulas could fill an encyclopedia. Brute force—harnessing the power of an ever growing number of computers—can fight off the tide of complexity for a while but soon succumbs to increasing numbers of external particles or loops.

Quantum fluctuations of space and time are much more innocuous than we imagined.

Even worse, what started as a concrete way to visualize the microscopic world can cloak it in obscurity. Individual Feynman diagrams are often impenetrably baroque, and when we have to juggle so many of them, we lose track of the essential physics. What is astounding is that the final result, found by summing up all the diagrams, can be quite simple. Different diagrams partially negate one another, and sometimes formulas with millions of terms collapse down to a single term. These cancellations suggest that the diagrams are the wrong tools for the job—like trying to pound in a nail with a feather. There must be a better way.

BEYOND FEYNMAN DIAGRAMS

OVER THE YEARS physicists tried out many new techniques to do calculations, each slightly better than the one before it, and gradually the outlines of an alternative to Feynman diagrams took shape. Our own involvement began in the early 1990s, when two of us (Bern and Kosower) showed how string theory could simplify QCD calculations by summarizing all the relevant Feynman diagrams with a single formula. With this formula, the three of us analyzed a particle reaction that had never been understood in detail before: the scattering of two gluons into three gluons, with one virtual-particle loop. This process was very complicated by the standards of the time but could be fully described by an amazingly simple formula, which fit on a single page.

The formula was so simple that, together with David Dunbar, who was at the University of California, Los Angeles, at the time, we found we could understand the scattering almost entirely in terms of a principle called unitarity. Unitarity is the requirement that the probabilities of all possible outcomes add up to 100 percent. (Technically the quantities are not probabilities but square roots of probabilities, but this distinction is not so important here.) Unitarity is implicit in Feynman's technique but tends to be hidden by the complexity of the calculations. So we developed an alternative technique that put it front and center. The idea of basing calculations on unitarity had come up in the 1960s but fell out of favor. As repeatedly happens in science, discarded ideas can come roaring back in a new guise.

The key to the success of the unitarity method is that it avoids the direct use of virtual particles, which are the prime reason that Feynman diagrams get so complicated. Such particles have both real and spurious effects. By definition, the latter *have* to cancel out of the final result, so they are excess mathematical baggage that physicists are happy to leave behind.

The method can be understood by analogy to a complicated subway system like the London Underground, with multiple paths between any two subway stations. Suppose we want to know the probability that a person entering at Mile End leaves at Wimbledon. The Feynman technique adds up the probabilities of all conceivable paths. All really means *all:* besides the paths through corridors and tunnels, Feynman diagrams include paths through solid rock where there are no subway lines or walkways. Those unrealistic paths are the analogues of the spurious contributions from virtual-particle loops. They cancel out in the end, but in the intermediate stages of the calculation, we need to keep track of them all.

In the unitarity approach, we consider only paths that make sense. We calculate the probability that a person takes a particular route by subdividing the problem: What is the probability the person goes through a particular turnstile, one pathway or another one, at each step of the journey? This procedure greatly cuts down the size of calculations.

The choice between the Feynman and unitarity methods is not a matter of right and wrong. Both express the same basic

© 2013 Scientific American

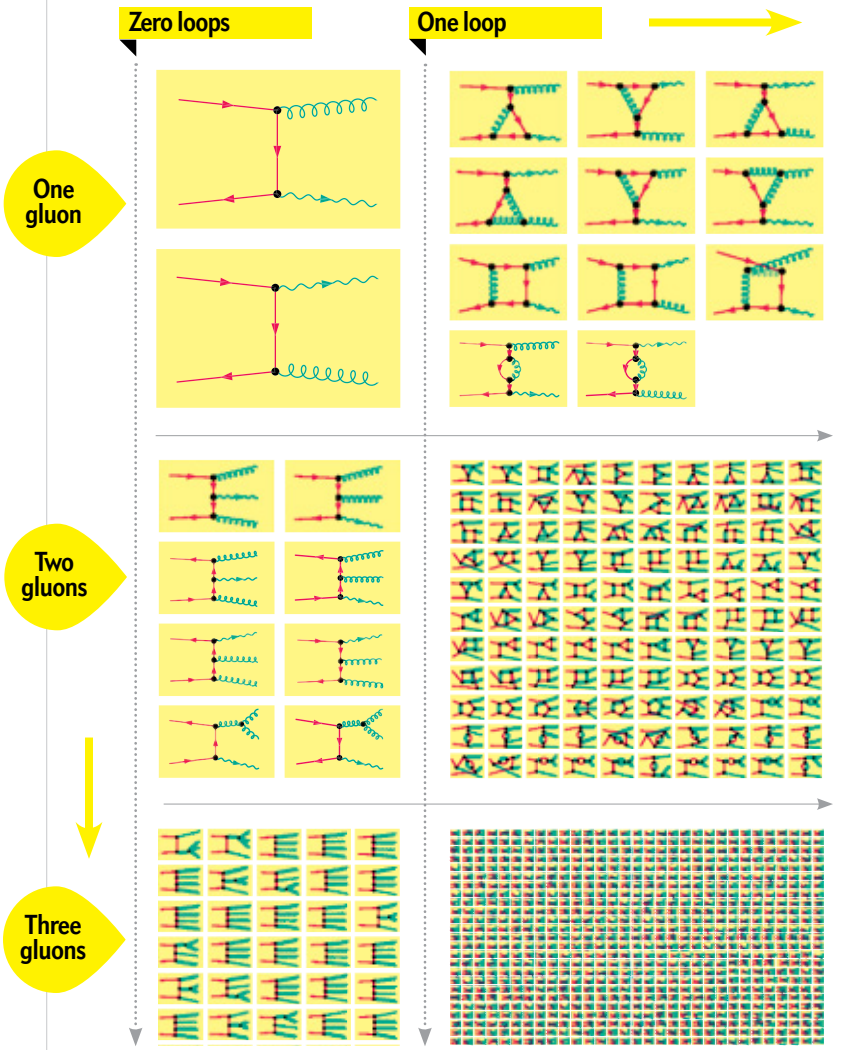
physical principles. Both would eventually lead to the same numerical probabilities. But they represent different levels of description. A single Feynman diagram, out of the tens of thousands for a messy collision, is like a single molecule within a droplet of fluid. In principle, you can determine what a fluid will do by tracking all the individual molecules, but that makes sense only for a microscopically small droplet. It is not only cumbersome but also unenlightening. The fluid could be cascading down a hill, but you would scarcely know that from the molecular description. You are better off considering higher-level properties such as fluid velocity, density and pressure. Likewise, instead of viewing a particle collision as built up one by one from individual Feynman diagrams, physicists can think of it holistically. We concentrate on the properties that govern the process as a whole—unitarity, as well as special symmetries to which the unitarity method gives prominence. In special cases, we can make theoretical predictions with perfect precision, which would take an infinite amount of time with Feynman diagrams.

The advantages go even further. After we developed the unitarity method for virtual-particle loops, another team, then at the Institute for Advanced Study in Princeton, N.J.—Ruth Britto, Freddy Cachazo, Bo Feng and Edward Witten—added a complementary twist. The theorists thought about tree diagrams again and computed the probability of a collision involving, say, five particles in terms of the probability of a collision of four particles, followed by one particle splitting into two. That is a stunning conclusion because a five-particle collision usually looks very different from these two sequential collisions. In more ways than one, we can subdivide daunting particle problems into simpler pieces.

WHY FEYNMAN DIAGRAMS DRIVE PHYSICISTS MAD

Too Many to Keep Track Of

Each Feynman diagram provides an intuitive way to visualize one possible way that particles might interact. The trouble is that there are countless other ways, too. A quark-quark interaction might produce more than one gluon or involve more than one virtual-particle loop, or both. The calculations quickly become unmanageable.



SMASHING WATCHES

PROTON COLLISIONS at the LHC are exceptionally complex. Feynman himself once compared such collisions to figuring out how Swiss watches work by smashing them together, and his technique struggles to track what goes on during these encounters. Protons are not elementary particles but little balls of quarks and gluons bound together by the strong subnuclear force. When they slam together, quarks can bounce off quarks, quarks off gluons, gluons off gluons. Quarks and gluons can split into still more quarks and gluons. They ultimately clump together in composite particles that shoot out of the collider in narrow sprays physicists call jets.

Somewhere buried in that mess may be things that human beings have never seen before: new particles, new symmetries, maybe even new dimensions of spacetime. But sifting them out will be tough. To our instruments, exotic particles look rather like ordinary ones. The differences are small and easily missed. With the unitarity method, we can describe ordinary physics so precisely that extraordinary physics will stand out.

For instance, Joe Incandela of U.C. Santa Barbara, who is currently the spokesperson for the 2,000-plus-physicist CMS experiment at the LHC, came to us with a question about his team's search for exotic particles that make up cosmic dark matter, the mysterious stuff that astronomers think is out there but that physicists have

© 2013 Scientific American

yet to identify. Any such particles the LHC produces would elude the CMS detector, leaving the impression that some energy had gone missing. Unfortunately, an apparent loss of energy does not, in itself, mean that the LHC has synthesized dark matter. For instance, the LHC frequently produces an ordinary particle called the *Z* boson, and one fifth of the time it decays into two neutrinos, which also interact very weakly and escape the detector without registering. How could one predict the number of Standard Model particles whose effects mimic dark particles?

Incandela's group proposed a solution: take the number of photons the CMS detector records, extrapolate to the number of events involving neutrinos and see whether they fully explain the apparent energy loss. If not, the LHC might be creating dark matter. This idea typifies the indirect estimates experimental physicists are always having to make because they lack the ability to observe certain types of particles directly. But to pull it off, Incandela's group needed to know precisely how the number of photons related to the number of neutrinos. Unless the precision was high enough, the backdoor strategy would fail. With several colleagues, we studied the problem using the new theoretical tools and were able to assure Incandela that the precision was sufficiently high.

This success has inspired us to push forward with more ambitious calculations. As is common in modern particle physics, we work with collaborators worldwide, including Fernando Febres Cordero of Simón Bolívar University in Caracas, Venezuela, Harald Ita, now at the University of Freiburg in Germany, Daniel Maître of Durham University in England, Stefan Höche of SLAC and Kemal Ozeren of U.C.L.A. Together we made precise predictions for the probability that LHC collisions would produce a pair of neutrinos, along with four jets. Using Feynman diagrams, these calculations would have been too imposing even for a large team of physicists working hard for a decade assisted by state-of-the-art computers. The unitarity method let us do them in less than a year. To our delight, another LHC team, the ATLAS collaboration, has already compared our predictions with its data, and the results are in excellent agreement. Going forward, experimenters will use these results to search for new physics.

The unitarity method has also aided the search for the recently discovered Higgs-like particle. A sign of the Higgs is the production of a single electron, a pair of jets and a neutrino, the neutrino again leaving the impression that energy has gone missing. The same outcome can also arise from particle reactions not involving the Higgs. One of our first uses of the unitarity method was to calculate the precise probability of these confounding reactions.

BACK TO GRAVITY

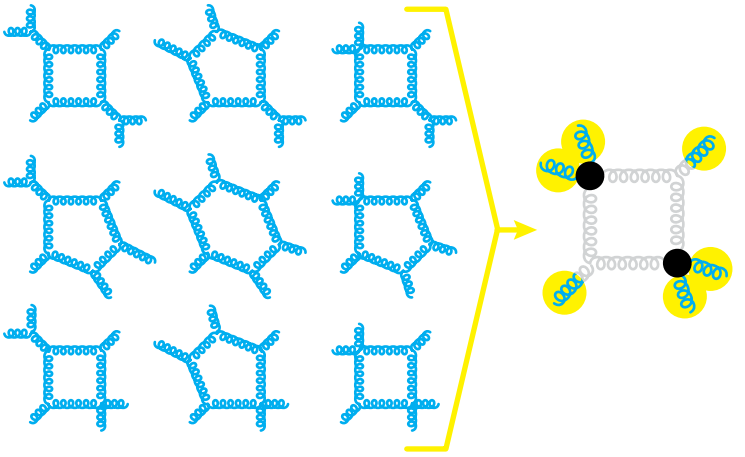
AN EVEN MORE IMPRESSIVE USE of the unitarity method is to study quantum gravity. For physicists to develop a fully consistent theory of nature, we must find a way to fit gravity into a quantum-mechanical framework. If gravity behaves like the other forces of nature, it should be transmitted by graviton particles. Gravitons would collide and scatter just as other particles do, and we could draw Feynman diagrams for them. Yet attempts in the mid-1980s to describe graviton scattering by quantizing Einstein's theory in the simplest possible way led to nonsensical predictions, such as infinite values for quantities that should clearly be finite. Infinite quantities, per se, are not the problem. They can arise at intermediate stages of calculations even in a well-behaved theory such as the Standard Model, but they should cancel out of any quantity that is potentially measurable. For gravity, no such cancellations appeared to occur. In concrete terms, it means that the quantum fluctuations of space and time, dubbed "spacetime foam" by the late quantum gravity pioneer John Wheeler, spiral out of control.

One possible explanation is that na-

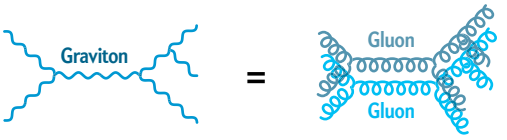
A BETTER WAY

Making Life Simpler

The authors have developed an alternative to Feynman diagrams called the unitarity method. It exploits basic principles in quantum theory to consolidate multiple Feynman diagrams, making calculations possible that were once intractable.



The technique has revealed features of the natural world that were implicit in current theories but got lost in the swarms of Feynman diagrams. Most striking among them is a possible way to incorporate the force of gravity, which has resisted past efforts to explain it. The particles that transmit the force of gravity, known as gravitons, bear a remarkable similarity to gluons; in fact, mathematically, each graviton is like two copies of a gluon, working together like contestants in a three-legged race.



EXTREME PHYSICS

© 2013 Scientific American

ture contains undiscovered particles that rein in these quantum effects. This idea, embodied in so-called supergravity theories, was studied intensively during the 1970s and early 1980s. But the excitement died down when indirect arguments suggested that nonsensical infinities would still arise with three or more virtual-particle loops. It seemed that supergravity was doomed to failure.

This disappointment led many to pursue string theory. String theory is a radical departure from the Standard Model. According to it, particles such as quarks, gluons and gravitons are no longer tiny points but oscillations of one-dimensional strings. Particle interactions are spread out over the strings rather than concentrated at a single point, preventing infinities automatically. On the other hand, string theory has encountered its own troubles; for instance, it does not make definitive theoretical predictions for observable phenomena.

DOUBLE TROUBLE

IN THE MID-1990S Stephen Hawking of the University of Cambridge advocated giving supergravity theories another look. He pointed out that the 1980s-era studies had taken shortcuts that made their conclusions questionable. But Hawking was unable to convince anyone else because there was a good reason that people had taken those shortcuts: the full calculations were hopelessly beyond the capacities of even the most brilliant math whiz. To know for sure whether a Feynman diagram with three virtual-graviton loops produces infinite quantities, we would need to evaluate 10^{20} terms. By five loops, a diagram spawns 10^{30} terms, roughly one for every atom in an LHC detector. The issue appeared destined for the dustbin of undecidable problems.

The unitarity method has completely changed the situation. Using it, we have conducted a physics version of the Innocence Project and reopened the case against supergravity theory. What would have taken the Feynman technique 10^{20} terms, we can now do with dozens. Together with Radu Roiban of Pennsylvania State University, as well as John Joseph Carrasco and Henrik Johansson, who were then graduate students at U.C.L.A., we found that the 1980s-era speculations were wrong. Quantities that seemed destined to be infinite are in fact finite. Supergravity is not as nonsensical as physicists thought. In concrete terms, it means that quantum fluctuations of space and time are much more innocuous in supergravity than previously imagined. If you ply us with fine wine, you might catch us speculating that some version of it might be the long-sought quantum theory of gravity.

Even more remarkably, three gravitons interact just like two copies of three interacting gluons. This double-copy property appears to persist no matter how many particles are scattering or how many virtual-particle loops are involved. It means that, figuratively speaking, gravity is the square of the strong subnuclear interaction. It will take us a while to translate the mathematics into physical insight and check whether it is true under all conditions. For now the crucial point is that gravity may not be so different from the other forces of nature.

As is common in science, after each debate is settled, another erupts. Immediately after our calculations for three loops, skeptics wondered whether trouble might appear at four loops. As also happens frequently, bottles of wine were bet on the outcome of the calculation: an Italian Barolo against a Napa Valley Chardonnay. When we did the calculation, we found no hints of difficulties, settling at least this debate (and popping a Barolo cork).

Is supergravity theory completely free of infinities? Or does its high degree of symmetry merely curb some of its excesses at a small number of loops? In the latter case, trouble should creep in at five loops; by seven loops, quantum effects should grow strong enough to produce infinities. David Gross of U.C. Santa Barbara has put up a bottle of California Zinfandel if no seven-loop infinities arise. To settle this latest bet, some of us have embarked on new calculations. An absence of seven-loop infinities would astound the skeptics and might finally persuade them that supergravity could be self-consistent. Even if it is, though, the theory does not capture other kinds of effects, called nonperturbative, that are too tiny to be seen in the loop-by-loop approach we have been following. Those effects may still require an even deeper theory to handle, perhaps string theory.

Physicists often like to think of new theories as emerging from the bold brushstrokes of new principles—relativity, quantum mechanics, symmetry. But sometimes those theories emerge from a careful reexamination of the principles we already know. The quiet revolution in our understanding of particle collisions has enabled us to work out the consequences of the Standard Model in remarkable detail, leading to significant improvements in our potential to discover physics beyond it. Even more surprisingly, it is letting us follow the unexplored implications of old physics, including a once neglected path toward unifying gravity with the other known forces. In many ways, the journey to understanding the secrets of how elementary particles scatter has not been like riding the predictable London Underground at all but more like a journey on the Knight Bus of Harry Potter tales, where you never quite know what will happen next. SA

Zvi Bern is a physics professor at the University of California, Los Angeles. In 2011 he co-organized a three-month workshop at the Kavli Institute for Theoretical Physics at U.C. Santa Barbara to discuss high-energy particle scattering. The eldest of his three children is a physics major at Brandeis University. The other two are still trying to decide whether their father's profession is cool.

Lance J. Dixon remembers getting interested in science when he watched the launch of Pioneer 10, the first space probe to visit the outer solar system. Today he is a professor at the SLAC National Accelerator Laboratory. He spent the 2010–2011 academic year at CERN near Geneva, talking with experimenters about how to best use the new theoretical predictions described in this article.

David A. Kosower is as passionate about science as he is about traveling, skiing and running. He is a senior scientist at the Institute of Theoretical Physics at the French Alternative Energies and Atomic Energy Commission's Saclay research center. He is a current holder of a European Research Council Advanced Grant.

MORE TO EXPLORE

Lovely as a Tree Amplitude: Hidden Structures Underlie Feynman Diagrams. Steven K. Blau in *Physics Today,* Vol. 57, No. 7, page 19; July 2004.

Physics and Feynman's Diagrams. David Kaiser in *American Scientist,* Vol. 93, No. 2, pages 156–165; March/April 2005.

Cancellations beyond Finiteness in N=8 Supergravity at Three Loops. Z. Bern, J. J. Carrasco, L. J. Dixon, H. Johansson, D. A. Kosower and R. Roiban in *Physical Review Letters,* Vol. 98, No. 16; April 20, 2007.

Supergravity: Finite after All? Kellogg Stelle in *Nature Physics,* Vol. 3, pages 448–450; July 2007.

Pulling QCD Predictions out of a (Black) Hat. Daniel Maître in *SLAC Today;* August 7, 2008.

A Simple Twist of Fate. George Musser in *Scientific American,* Vol. 302, No. 6, pages 14–17; June 2010.

Precise Predictions for W+4 Jet Production at the Large Hadron Collider. C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower and D. Maître in *Physical Review Letters,* Vol. 106, No. 9; March 1, 2011.

© 2013 Scientific American

© 2013 Scientific American

COSMOLOGY

ORIGIN OF THE UNIVERSE

Cosmologists are closing in on the ultimate processes that created and shaped the universe

By Michael S. Turner

THE UNIVERSE IS BIG IN BOTH SPACE AND TIME AND, FOR MUCH OF humankind's history, was beyond the reach of our instruments and our minds. That changed dramatically in the 20th century. The advances were driven equally by powerful ideas—from Einstein's general relativity to modern theories of the elementary particles—and powerful instruments—from the 100- and 200-inch reflectors that George Ellery Hale built, which took us beyond our Milky Way galaxy, to the Hubble Space Telescope, which has taken us back to the birth of galaxies. Over the past 30 years the pace of progress has accelerated with the realization that dark matter is not made of ordinary atoms, the discovery of dark energy, and the dawning of bold ideas such as cosmic inflation and the multiverse.

IN BRIEF

Our universe began with a hot big bang 13.7 billion years ago and has expanded and cooled ever since. It has evolved from a formless soup of elementary particles into the richly structured cosmos of today.

The first microsecond was the formative period when matter came to dominate over antimatter, the seeds for galaxies and other structures were planted, and dark matter (the unidentified material that holds those structures together) was created.

The universe's future lies in the hands of dark energy, an unknown form of energy that caused cosmic expansion to start accelerating a few billion years ago.

GENE BURKHARDT (*styling*)

Photoillustration by Holly Lindem

© 2013 Scientific American

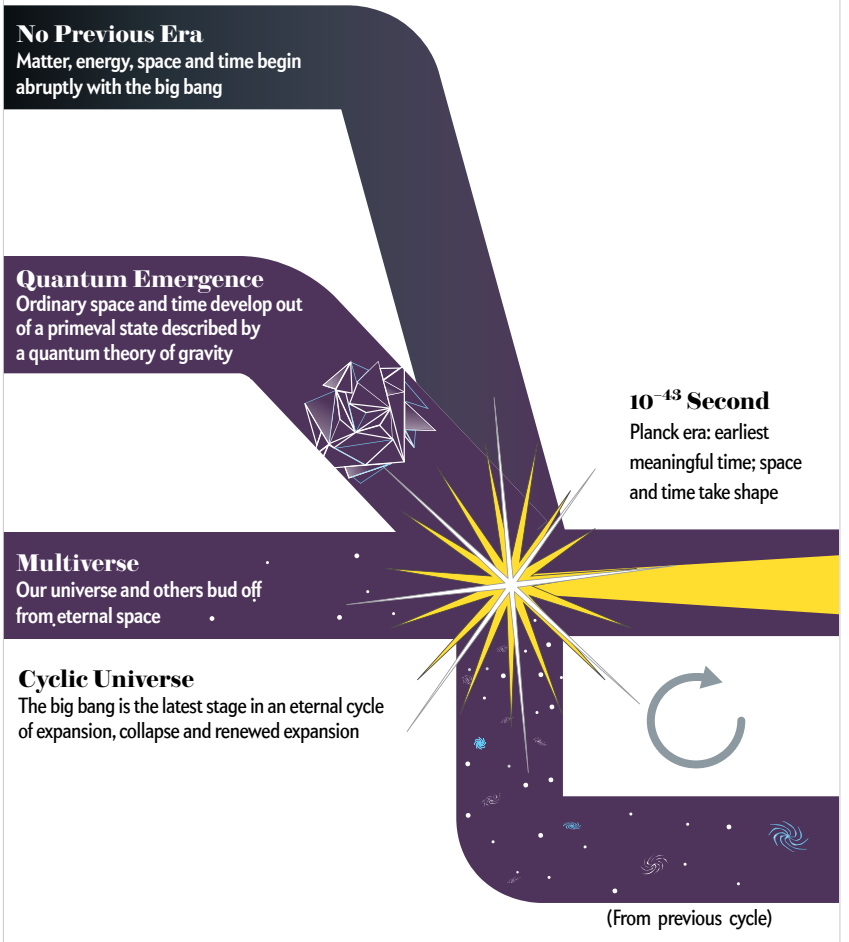
The universe of 100 years ago was simple: eternal, unchanging, consisting of a single galaxy, containing a few million visible stars. The picture today is more complete and much richer. The cosmos began 13.7 billion years ago with the big bang. A fraction of a second after the beginning, the universe was a hot, formless soup of the most elementary particles, quarks and leptons. As it expanded and cooled, layer on layer of structure developed: neutrons and protons, atomic nuclei, atoms, stars, galaxies, clusters of galaxies, and finally superclusters. The observable part of the universe is now inhabited by 100 billion galaxies, each containing 100 billion stars and probably a similar number of planets. Galaxies themselves are held together by the gravity of the mysterious dark matter. The universe continues to expand and indeed does so at an accelerating pace, driven by dark energy, an even more mysterious form of energy whose gravitational force repels rather than attracts.

The overarching theme in our universe's story is the evolution from the simplicity of the quark soup to the complexity we see today in galaxies, stars, planets and life. These features emerged one by one over billions of years, guided by the basic laws of physics. In our journey back to the beginning of creation, cosmologists first travel through the well-established history of the universe back to the first microsecond; then to within 10^{-34} second of the beginning, for which ideas are well formed but the evidence is not yet firm; and finally to the earliest moments of creation, for which our ideas are still just speculation. Although the ultimate origin of the universe still lies beyond our grasp, we have tantalizing conjectures, including the notion of the multiverse, whereby the universe comprises an infinite number of disconnected subuniverses.

COSMIC TIMELINE

Before the Big Bang

Cosmologists do not yet know how the universe began, but this question has now come within the realm of science, with a number of speculative scenarios being discussed.


No Previous Era
Matter, energy, space and time begin abruptly with the big bang
Quantum Emergence
Ordinary space and time develop out of a primeval state described by a quantum theory of gravity
10^{-43} Second
Planck era: earliest meaningful time; space and time take shape
Multiverse
Our universe and others bud off from eternal space
Cyclic Universe
The big bang is the latest stage in an eternal cycle of expansion, collapse and renewed expansion
(From previous cycle)

EXPANDING UNIVERSE

USING THE 100-INCH HOOKER TELESCOPE on Mount Wilson in 1924, Edwin Hubble showed that fuzzy nebulae, studied and speculated about for several hundred years, were galaxies just like our own—thereby enlarging the known universe by 100 billion. A few years later he showed that galaxies are moving apart from one another in a regular pattern described by a mathematical relation now known as Hubble's law, according to which galaxies that are farther away are moving faster. It is Hubble's law, played back in time, that points to a big bang 13.7 billion years ago.

Hubble's law found ready interpretation within general relativity: space itself is expanding, and galaxies are being carried along for the ride [*see box on opposite page*]. Light, too, is being stretched, or redshifted—a process that saps its energy, so that the universe cools as it expands. Cosmic expansion provides the narrative for understanding how today's universe came to be. As cosmologists imagine rewinding the clock, the universe becomes denser, hotter, more extreme and simpler. In exploring the beginning, we also probe the inner workings of nature by taking advantage of an accelerator more powerful than any built on Earth—the big bang itself.

By looking out into space with telescopes, astronomers peer back in time—and the larger the telescope, the farther back they peer. The light from distant galaxies reveals an earlier epoch, and the amount this light has redshifted indicates how much the universe has grown in the intervening years. The current record holder has a redshift of more than 10, representing a time when the universe was less than one-eleventh its present size and only a few hundred million years old. Telescopes such as the Hubble Space Telescope and the 10-meter Keck telescopes on Mauna Kea rou-

Illustrations by Melissa Thomas

© 2013 Scientific American

tinely take us back to the epoch when galaxies like ours were forming, a few billion years after the big bang. Light from even earlier times is so strongly redshifted that astronomers must look for it in the infrared and radio bands. Telescopes such as the planned James Webb Space Telescope, a 6.5-meter infrared telescope, and the Atacama Large Millimeter Array (ALMA), a network of 66 radio dishes already operating in northern Chile, can take us back to the birth of the very first stars and galaxies.

Computer simulations say that those stars and galaxies emerged when the universe was about 100 million years old. Before then, the universe went through a time called the "dark ages," when it was almost pitch-black. Space was filled with a featureless gruel, five parts dark matter and one part hydrogen and helium, that thinned out as the universe expanded. Matter was slightly uneven in density, and gravity acted to amplify these density variations: denser regions expanded more slowly than less dense ones did. By 100 million years the densest regions did not merely expand more slowly but actually started to collapse. Such regions contained about one million solar masses of material each. They were the first gravitationally bound objects in the cosmos.

Dark matter accounted for the bulk of their mass but was, as its name suggests, unable to emit or absorb light. So it remained in an extended cloud. Hydrogen and helium gas, on the other hand, emitted light, lost energy and became concentrated in the center of the cloud. Eventually it collapsed all the way down to stars. These first stars were much more massive than today's—hundreds of solar masses. They lived very short lives before exploding and leaving behind the first heavy elements. Over the next billion years or so the force of gravity assembled these million-solar-mass clouds into the first galaxies.

Radiation emitted from primordial hydrogen clouds, which were greatly redshifted by the expansion, should be detectable by giant arrays of radio antennas with a total collecting area of up to one square kilometer. When built, these arrays will watch as the first generation of stars and galaxies ionize the hydrogen and bring the dark ages to an end.

HUBBLE ULTRA DEEP FIELD, the most sensitive optical image of the cosmos ever made, reveals more than 1,000 galaxies in their early stages of formation.

COURTESY OF R. THOMPSON (*University of Arizona*, NASA AND ESA (*Hubble Ultra Deep Field*)

FAINT GLOW OF A HOT BEGINNING

BEYOND THE DARK AGES is the glow of the hot big bang at a redshift of 1,100. This radiation has been redshifted from visible light (a red-orange glow) beyond even the infrared to microwaves. What we see from that time is a wall of microwave radiation filling the sky—the cosmic microwave background radiation (CMB), discovered in 1964 by Arno Penzias and Robert Wilson. It provides a glimpse of the universe at the tender age of 380,000 years, the period when atoms formed. Before then, the universe was a nearly uniform soup of atomic nuclei, electrons and photons. As it cooled to a temperature of about 3,000 kelvins, the nuclei and electrons came together to form atoms. Photons ceased to scatter off electrons and streamed across space unhindered, revealing the universe at a simpler time before the existence of stars and galaxies.

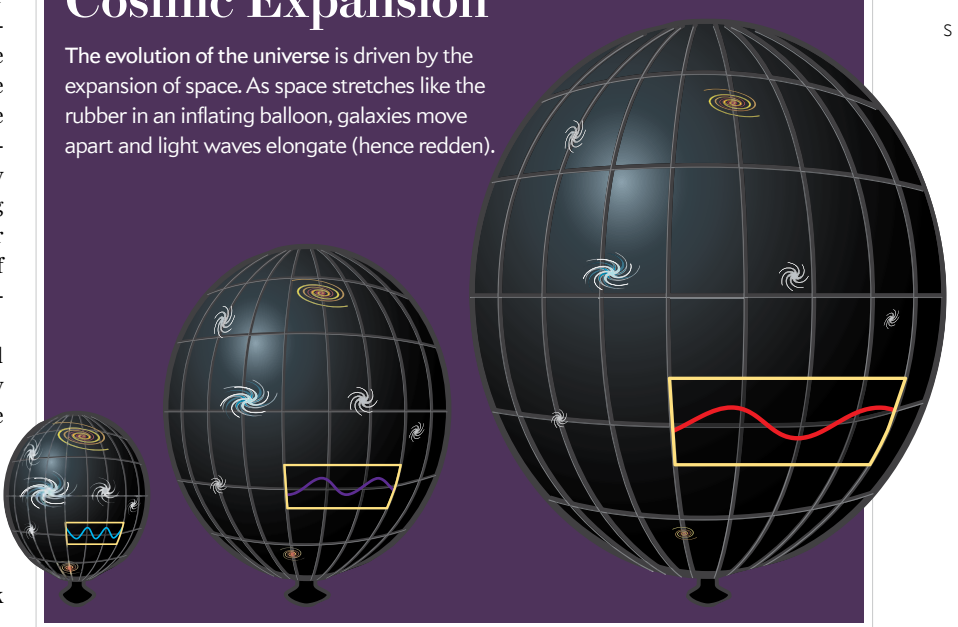
In 1992 NASA's Cosmic Background Explorer satellite discovered that the intensity of the CMB has slight variations—about 0.001 percent—reflecting a slight lumpiness in the distribution of matter. The degree of primordial lumpiness was enough to act as seeds for the galaxies and larger structures that would later emerge from the action of gravity. The pattern of these variations in the CMB across the sky also encodes basic properties of the universe, such as its overall density and composition, as well as hints about its earliest moments; the careful study of these variations has revealed much about the universe [*see illustration on page 41*].

As we roll a movie of the universe's evolution back from that point, we see

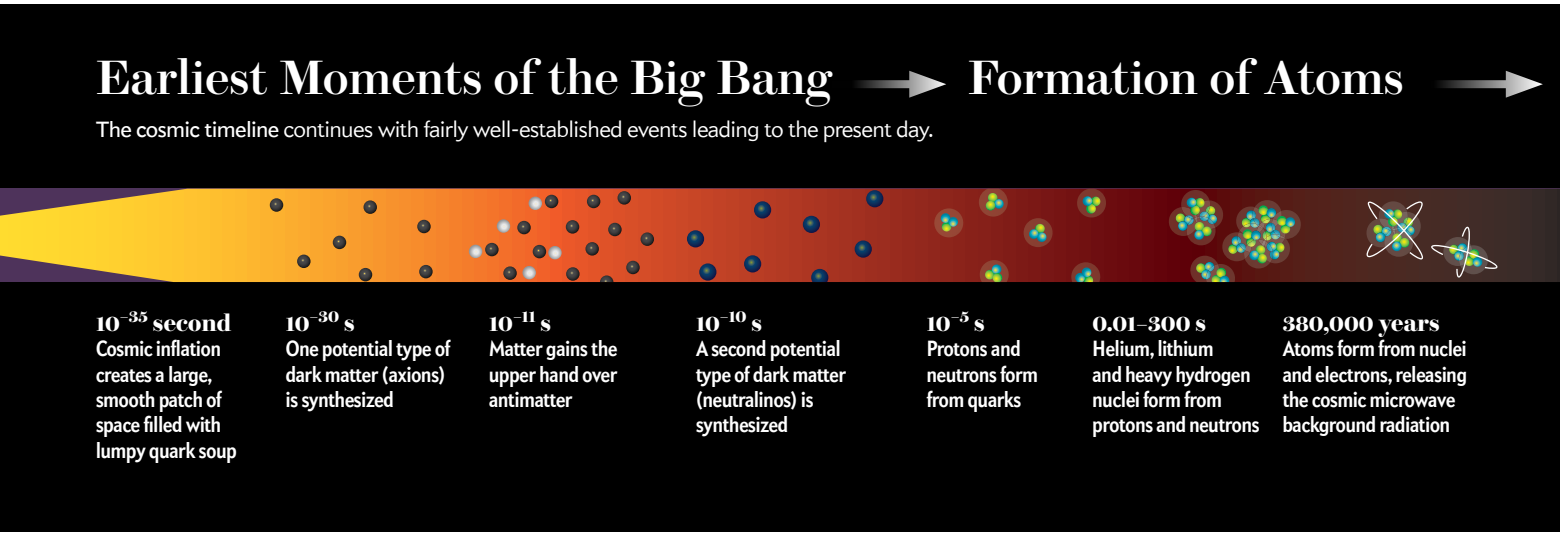
PRIMER

Cosmic Expansion

The evolution of the universe is driven by the expansion of space. As space stretches like the rubber in an inflating balloon, galaxies move apart and light waves elongate (hence redden).



© 2013 Scientific American

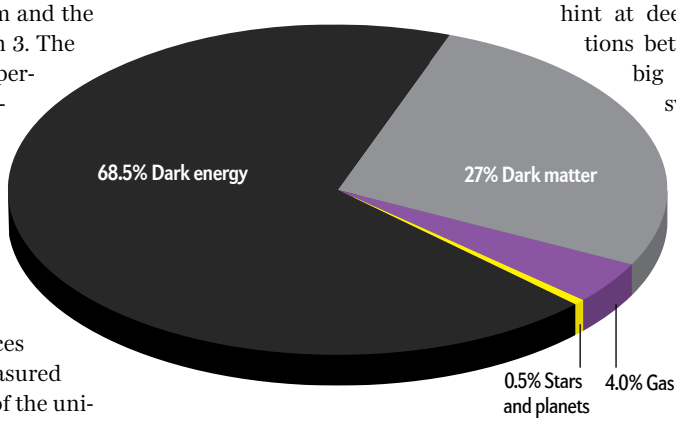

Earliest Moments of the Big Bang
Formation of Atoms
The cosmic timeline continues with fairly well-established events leading to the present day.
10^{-35} second
Cosmic inflation creates a large, smooth patch of space filled with lumpy quark soup
10^{-30} s
One potential type of dark matter (axions) is synthesized
10^{-11} s
Matter gains the upper hand over antimatter
10^{-10} s
A second potential type of dark matter (neutralinos) is synthesized
10^{-5} s
Protons and neutrons form from quarks
0.01–300 s
Helium, lithium and heavy hydrogen nuclei form from protons and neutrons
380,000 years
Atoms form from nuclei and electrons, releasing the cosmic microwave background radiation

the primordial plasma becoming ever hotter and denser. Prior to about 100,000 years, the energy density of radiation exceeded that of matter, which kept matter from clumping. Therefore, this time marks the beginning of gravitational assembly of all the structure seen in the universe today. Still further back, when the universe was less than a second old, atomic nuclei had yet to form; only their constituent particles—namely, protons and neutrons—existed. Nuclei emerged when the universe was seconds old and the temperatures and densities were just right for nuclear reactions. This process of big bang nucleosynthesis produced only the lightest elements in the periodic table: a lot of helium (about 25 percent of the atoms in the universe by mass) and smaller amounts of lithium and the isotopes deuterium and helium 3. The rest of the plasma (about 75 percent) stayed in the form of protons that would eventually become hydrogen atoms. All the rest of the elements in the periodic table formed billions of years later in stars and stellar explosions.

Nucleosynthesis theory accurately predicts the abundances of elements and isotopes measured in the most primeval samples of the universe—namely, the oldest stars and high-redshift gas clouds. The abundance of deuterium, which is very sensitive to the density of atoms in the universe, plays a special role: its measured value implies that ordinary matter amounts to 4.5 ± 0.1 percent of the total energy density. (The remainder is dark matter and dark energy.) This estimate agrees precisely with the composition that has been gleaned from the analysis of the CMB. This correspondence is a great triumph. That these two very different measures, one based on nuclear physics when the universe was a second old and the other based on atomic physics when the universe was 380,000 years old, agree is a strong check not just on our model of how the cosmos evolved but on all of modern physics.

ANSWERS IN THE QUARK SOUP

EARLIER THAN A MICROSECOND, even protons and neutrons could not exist and the universe was a soup of nature's basic building blocks: quarks, leptons, and the force carriers (photons, the *W* and *Z* bosons, and gluons). We can be confident that the quark soup existed because experiments at particle accelerators have re-created similar conditions here on Earth today.

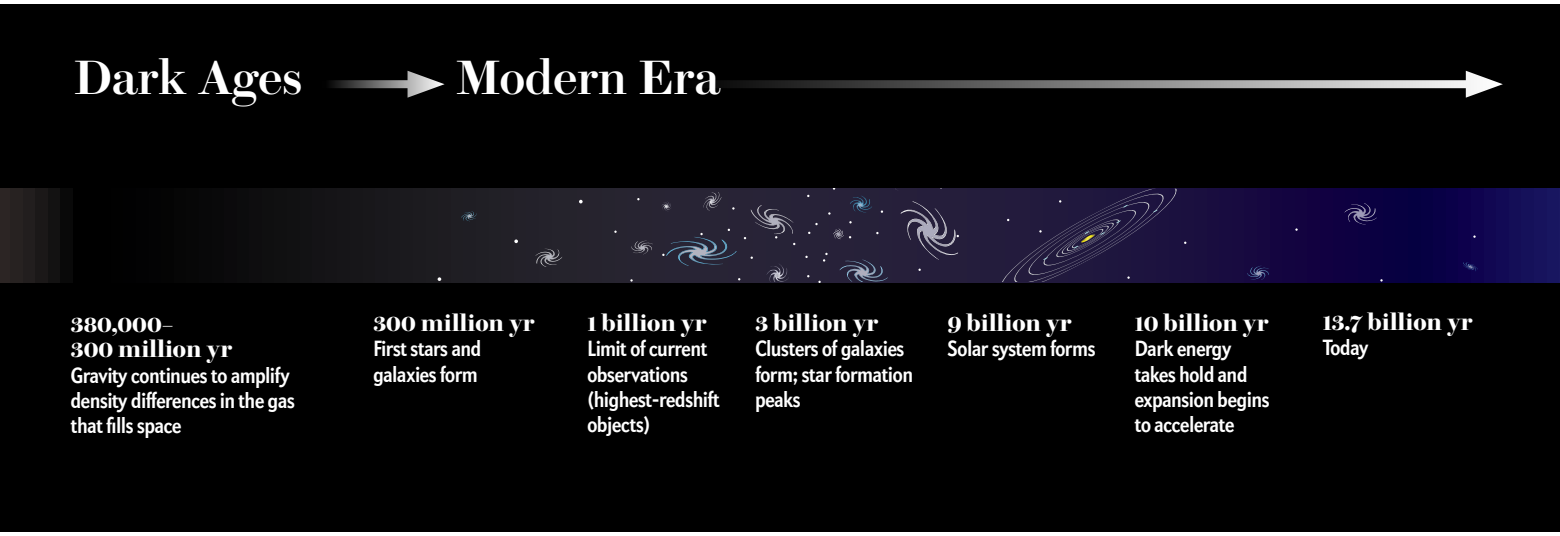


BULK OF UNIVERSE consists of dark energy and dark matter, neither of which has been identified. Ordinary matter of the kind that makes up stars, planets and interstellar gas accounts for only a small fraction.

To explore this epoch, cosmologists rely not on bigger and better telescopes but also on powerful ideas from particle physics. The development of the Standard Model of particle physics 30 years ago has led to bold speculations, including string theory, about how the seemingly disparate fundamental particles and forces are unified. As it turns out, these new ideas have implications for cosmology that are as important as the original idea of the hot big bang. They hint at deep and unexpected connections between the world of the very big and of the very small. Answers to three key questions—the nature of dark matter, the asymmetry between matter and antimatter, and the origin of the lumpy quark soup itself—have been starting to emerge.

It now appears that the early quark soup phase was the birthplace of dark matter. The identity of dark matter remains unclear, but its existence is very well established. Our galaxy and every other galaxy, as well as clusters of galaxies, are held together by the gravity of unseen dark matter. Whatever the

© 2013 Scientific American

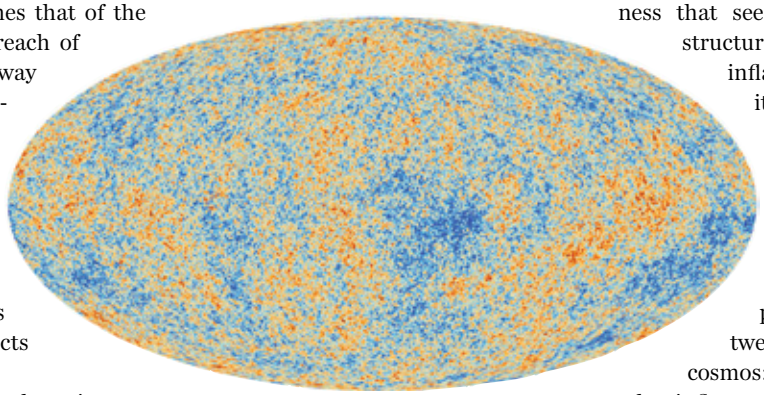

Dark Ages
Modern Era
380,000–
300 million yr
Gravity continues to amplify density differences in the gas that fills space
300 million yr
First stars and galaxies form
1 billion yr
Limit of current observations (highest-redshift objects)
3 billion yr
Clusters of galaxies form; star formation peaks
9 billion yr
Solar system forms
10 billion yr
Dark energy takes hold and expansion begins to accelerate
13.7 billion yr
Today

dark matter is, it must interact weakly with ordinary matter; otherwise it would have shown itself in other ways. Attempts to find a unifying framework for the forces and particles of nature have led to the prediction of stable or long-lived particles that might constitute dark matter. Some of these hypothetical particles would be present today as remnants of the quark soup phase in the correct numbers to be the dark matter and could even be detected.

One candidate is the called the neutralino, the lightest of a putative new class of particles that are heavier counterparts of the known particles. The neutralino is thought to have a mass between 100 and 1,000 times that of the proton, just within the reach of experiments now under way at the Large Hadron Collider at CERN near Geneva. Physicists have also built ultrasensitive underground detectors, as well as satellite and balloon-borne varieties, to look for this particle or the by-products of its interactions.

A second candidate is the axion, a superlightweight particle about one-trillionth the mass of the electron. Its existence is hinted at by subtleties that the Standard Model predicts in the behavior of quarks. Efforts to detect it exploit the fact that in a very strong magnetic field, an axion can transform into a photon. Both neutralinos and axions have the important property that they are, in a specific technical sense, "cold." Although they formed under broiling hot conditions, they were slow-moving and thus easily clumped into galaxies.

The early quark soup phase probably also holds the secret to why the universe today contains mostly matter rather than both matter and antimatter. Physicists think the universe originally had equal amounts of each, but at some point it developed a slight excess of matter—about one extra quark for every billion antiquarks. This imbalance ensured that enough quarks would survive annihilation with antiquarks as the universe expanded and cooled. More than 40 years ago accelerator experiments revealed that the laws of physics are ever so slightly biased in favor of matter, and in a still to be understood series of particle interactions very early on, this slight bias led to the creation of the quark excess.



COURTESY OF ESA/PLANCK COLLABORATION (*cosmic microwave radiation*)

COSMIC MICROWAVE background radiation is a snapshot of the universe at the tender age of 380,000 years. Tiny variations in the intensity of the radiation (*color-coded here*) are a cosmic Rosetta Stone that reveals key features of the universe, including its age, density, geometry and overall composition.

The quark soup itself is thought to have arisen at an extremely early time—perhaps 10^{-34} second after the big bang in a burst of cosmic expansion known as inflation. This burst, driven by the energy of a new field (thought to be distantly related to the recently discovered Higgs field) called the inflaton, would explain such basic properties of the cosmos as its general uniformity and the lumpiness that seeded galaxies and other structures in the universe. As the inflaton field decayed away, it released its remaining energy into quarks and other particles, thereby creating the heat of the big bang and the quark soup itself.

Inflation leads to a profound connection between the quarks and the cosmos: quantum fluctuations in the inflaton field on the subatomic scale get blown up to astrophysical size by the rapid expansion and become the seeds for all the structure we see today. In other words, the pattern seen on the CMB sky is a giant image of the subatomic world. Observations of the CMB agree with this prediction, providing the strongest evidence that inflation or

© 2013 Scientific American

something like it occurred very early in the history of the universe.

BIRTH OF THE UNIVERSE

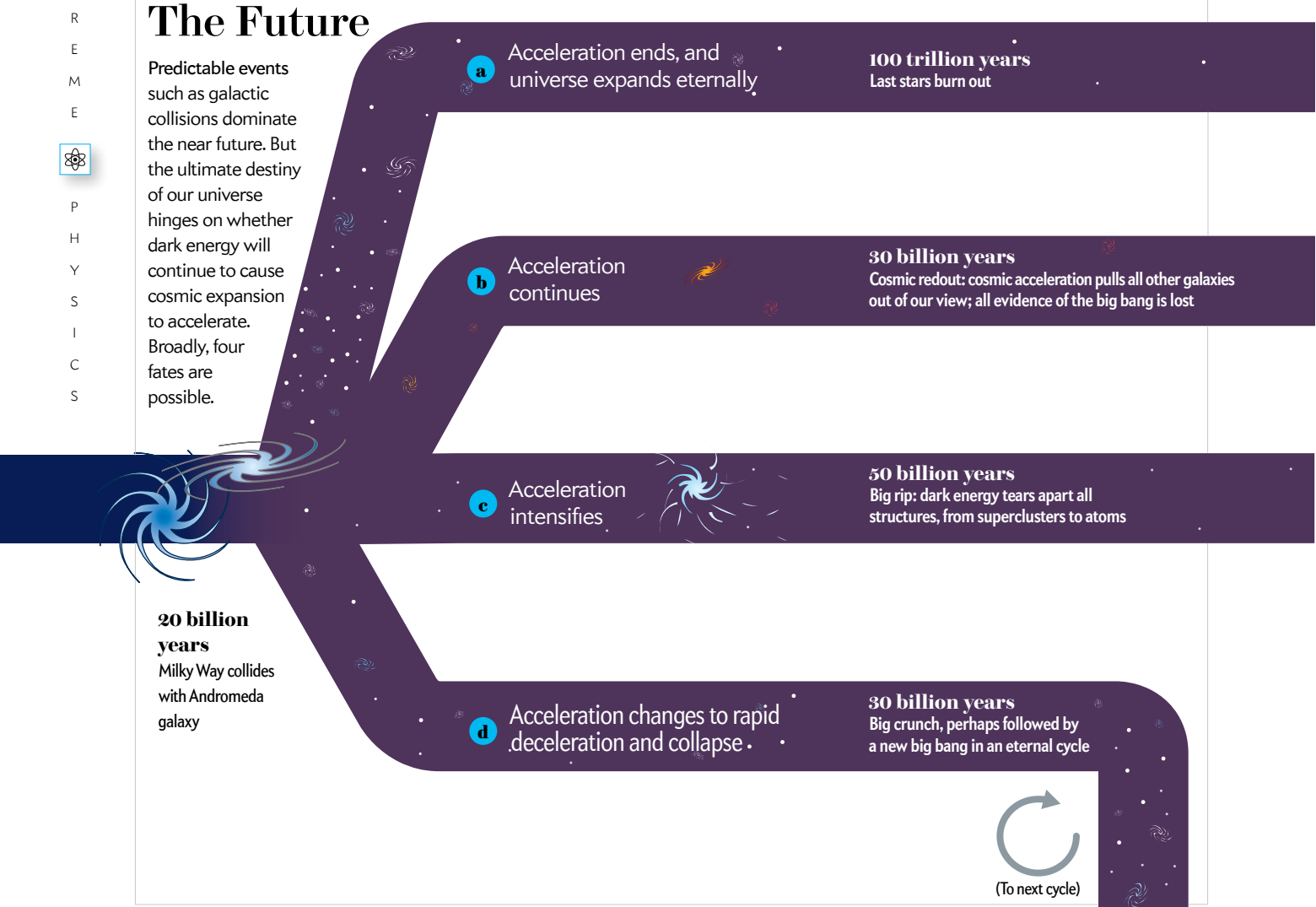
AS COSMOLOGISTS TRY to go even further to understand the beginning of the universe itself, our ideas become less firm. Einstein's general theory of relativity has provided the theoretical foundation for a century of progress in our understanding of the evolution of the universe. Because the general theory of relativity does not incorporate quantum theory, the other pillar of contemporary physics, it cannot be relied upon to address the very earliest moments of creation when quantum gravity effects should have been important. The discipline's greatest challenge is to develop a quantum theory of gravity, with which we will be able to address the so-called Planck era prior to about 10^{-43} second, when spacetime itself was taking shape.

Tentative attempts at a unified theory have led to some remarkable speculations about our very beginnings. String theory, for example, predicts the existence of additional dimensions of space and possibly other universes floating in that larger space. What we call the big bang may have been the collision of our universe with another. The marriage of string theory with the concept of inflation has led to perhaps the boldest idea yet, that of a multiverse—namely, that the universe comprises an infinite number of disconnected pieces, each with its own local laws of physics.

The multiverse concept, which is still in its infancy, turns on two key theoretical findings. First, the equations describing inflation strongly suggest that if inflation happened once, it should happen again and again, with an infinite number of inflationary regions created over time. Nothing can travel between these regions, so they have no effect on one another. Second, string theory suggests that these regions have different physical parameters, such as the number of spatial dimensions and the kinds of stable particles.

The idea of the multiverse provides novel answers to two of the biggest questions in all of science: what happened before the big bang and why the laws of physics are as they are (Albert Einstein's

COSMIC TIMELINE


The Future
Predictable events such as galactic collisions dominate the near future. But the ultimate destiny of our universe hinges on whether dark energy will continue to cause cosmic expansion to accelerate. Broadly, four fates are possible.
a
Acceleration ends, and universe expands eternally
100 trillion years
Last stars burn out
b
Acceleration continues
30 billion years
Cosmic redout: cosmic acceleration pulls all other galaxies out of our view; all evidence of the big bang is lost
c
Acceleration intensifies
50 billion years
Big rip: dark energy tears apart all structures, from superclusters to atoms
20 billion years
Milky Way collides with Andromeda galaxy
d
Acceleration changes to rapid deceleration and collapse
30 billion years
Big crunch, perhaps followed by a new big bang in an eternal cycle
(To next cycle)

© 2013 Scientific American

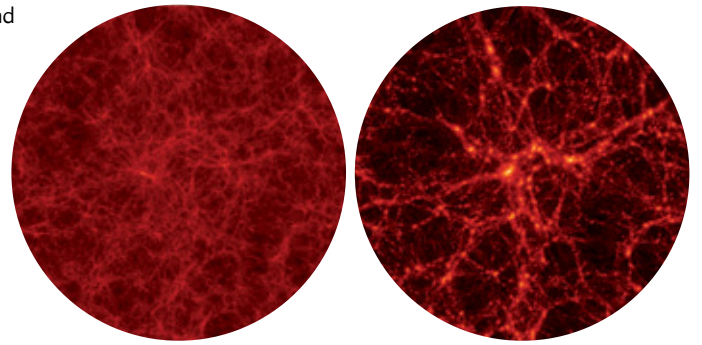
COSMIC ACCELERATION AND DARK ENERGY

In the Dark

A central feature of our current view of the universe, as well as its biggest mystery, is dark energy, the recently discovered and very weird form of energy that is causing cosmic expansion to speed up. Dark energy took control from matter a few billion years ago. Prior to that, the expansion had been slowing because of the gravitational attraction exerted by matter, and gravity was able to forge structures from galaxies to superclusters. Now, because of the influence of dark energy, structures larger than superclusters cannot form. In fact, had dark energy taken over earlier than it did—say, when the universe was only 100 million years old—structure formation would have ceased before even galaxies could have developed, and we would not be here.

Cosmologists have only rudimentary clues as to what dark energy might be. To speed up expansion of the universe requires a repulsive force, and Einstein's general theory of relativity predicts that the gravity of an extremely elastic form of energy can be actually be repulsive. The quantum energy that fills empty space acts in this way. The trouble is that theoretical estimates of the amount of quantum vacuum energy do not match the amount required by observations of the universe's expansion; in fact, they exceed it by many orders of magnitude. Alternatively, cosmic acceleration might be driven not by a new type of energy but by a process that mimics the repulsive force of such energy, perhaps the breakdown of general relativity or the influence of unseen spatial dimensions.

If the universe continues to accelerate at the current rate, then in 30 billion years all traces of the big bang will disappear. The light from all but a handful of nearby galaxies will be too redshifted to detect; the temperature of the cosmic background radiation will be too low to measure; and the universe will appear similar to one that astronomers knew 100 years ago before their instruments were powerful enough to reveal the universe we know today. —*M.S.T.*



IF THE UNIVERSE had even more dark energy than it does, it would have remained almost formless (*left*), without the large structures that we see (*right*).

famous musing about "whether God had any choice" about the laws). The multiverse makes moot the question of what happened before the big bang because there were an infinite number of big bang beginnings, each triggered by its own burst of inflation. Likewise, Einstein's question is pushed aside: within the infinity of universes, all possibilities for the laws of physics have been tried, so there is no particular reason for the laws that govern our universe.

Cosmologists have mixed feelings about the multiverse. If the disconnected subuniverses are truly incommunicado, we cannot hope to test their existence; they seem to lie beyond the realm of science. Part of me wants to scream, One universe at a time, please! On the other hand, the multiverse solves various conceptual problems. If correct, it will make Hubble's enlargement of the universe by a mere factor of 100 billion and Copernicus's banishment of Earth from the center of the universe in the 16th century seem like small advances in the understanding of our place in the cosmos.

Modern cosmology has humbled us. We are made of protons, neutrons and electrons, which together account for only 4.5 percent of the universe, and we exist only because of subtle connections between the very small and the very large. Events guided by the microscopic laws of physics allowed matter to dominate over antimatter, generated the lumpiness that seeded galaxies, filled space with dark matter particles that provide the gravitational infrastructure, and ensured that dark matter could build galaxies before dark energy became significant and the expansion began to accelerate [*see box above*]. At the same time, cosmology by its very nature is arrogant. The idea that we can understand something as vast in both space and time as our universe is, on the face of it, preposterous. This strange mix of humility and arrogance has gotten us pretty far in the past century in advancing our understanding of the present universe and its origin. I am bullish on further progress in the coming years, and I firmly believe we are living in a golden age of cosmology. SA

Michael S. Turner pioneered the interdisciplinary union of particle physics, astrophysics and cosmology with his seminal contributions to inflationary theory and cold dark matter theory and his prediction of dark energy (which he also named). He is director of the Kavli Institute for Cosmological Physics and Bruce V. and Diana M. Rauner Distinguished Service Professor at the University of Chicago. He is also president of the American Physical Society, a member of the National Academy of Sciences, and the recipient of the Helen B. Warner Prize for Astronomy and Dannie Heineman Prize for Astrophysics of the American Astronomical Society and the Julius Edgar Lilienfeld Prize of the American Physical Society.

MORE TO EXPLORE

The Early Universe. Edward W. Kolb and Michael S. Turner. Westview Press, 1994.
The Inflationary Universe. Alan Guth. Basic, 1998.
Quarks and the Cosmos. Michael S. Turner in *Science*, Vol. 315, pages 59–61; January 5, 2007.
Dark Energy and the Accelerating Universe. Joshua Frieman, Michael S. Turner and Dragan Huterer in *Annual Reviews of Astronomy and Astrophysics*, Vol. 46, pages 385–432; 2008. http://arxiv.org/abs/0803.0982

COURTESY OF LIANG GAO, CARLOS FRENK AND ADRIAN JENKINS *Institute for Computational Cosmology, Durham University*

© 2013 Scientific American

HIGHEST-ENERGY SUPERNOVAE might look quite spectacular from a planet orbiting the exploding star, but any civilization would most likely be obliterated.

© 2013 Scientific American

ASTROPHYSICS

SUPER SUPERNOVAE

The largest stars die in explosions more powerful than anyone thought possible—some triggered in part by the production of antimatter

By Avishay Gal-Yam

IN THE MIDDLE OF 2005 THE W. M. KECK OBSERVATORY ON MAUNA KEA IN Hawaii completed an upgrade of one of its giant twin telescopes. By automatically correcting for atmospheric turbulence, the instrument could now produce images as sharp as those from the Hubble Space Telescope. Shrinivas Kulkarni of the California Institute of Technology urged young Caltech researchers—myself among them—to apply for observing time. Once the rest of the astronomy community realized how terrific the telescopes were, he warned us, securing a slot would become very competitive.

Taking this advice, I teamed up with my then fellow postdocs Derek Fox and Doug Leonard to attempt a type of study that previously had been carried out almost solely with the Hubble: hunting for supernova progenitors. In other words, we wanted to know what stars look like when they are about to explode.

For decades theorists have been able to predict which celestial bodies are going to go supernova—for instance, they know that bright blue stars are due to explode soon. But "soon" to an astronomer means within the next million years or so. Thus, just patiently watching an individual star was not an option.

We thought that Keck could help us, and we were granted a single night of observing time in November 2005. As I flew in to the Big Island, I was anxious, hoping for good weather, as we had only one chance to try this new approach. Fortunately, the weather gods cooperated. That evening of observing set me on a research path that ultimately helped to overturn long-standing views of how large stars can become and how these giants die.

At the time, experts maintained that very large stars do not explode; rather they gradually shrink by shedding mass as stellar wind. Indeed, most theoretical astrophysicists would have said that because of these powerful winds, stars in the present-day universe cannot grow to a massive size in the first place—that they cannot become much heavier than, say, 100 times the mass of our sun.

As a result of our Hawaiian adventure, though, we gradually came to realize that stars of at least 200 solar masses do exist in our current universe and that they end their lives with the most energetic explosions in the universe. Equally surprising, we were also to discover that some of those stars explode in a way quite unlike anything astronomers had ever seen—in a process involving the generation of antimatter at the star's center.

Such enormous stars, and probably even larger ones, were the first celestial bodies to form from primordial gas in the universe's early history. Their way of exploding thus tells us how the elements

Illustration by Ron Miller

© 2013 Scientific American

they produced could spread around the cosmos and ultimately sow the seeds of today's suns, planets and people.

AN UNLIKELY START

IN OUR ONE TIME at the telescope, Fox, Leonard and I hoped to observe an active supernova and then, by looking at archival images shot by the Hubble, find an image of the star before it exploded. We therefore needed to look for a supernova in one of the many galaxies the Hubble had imaged in the past. The difficult part of finding our target in a Hubble photograph would be figuring out which star, among the billions in a galaxy, was the one that blew up. To do so, we would need to measure the location of the supernova with great precision. Before the advent of adaptive-optics systems such as Keck's, that was possible only through the Hubble itself. Even then, the task was so challenging that astronomers had managed to positively identify only three progenitors.

Among the supernovae active at the time, we selected one named SN 2005gl. Other groups would have considered it a poor choice—and for good reason: researchers who seek supernovae progenitors typically look within a radius of about 60 million light-years of Earth; this one was more than three times farther than that—about 200 million light-years away. For us to find the progenitor of SN 2005gl in Hubble images, that star would have to have been among the most luminous ever observed. The likelihood of success was low, but we felt that sometimes only by aiming at long shots can you reap huge rewards.

Our gamble paid off. After measuring SN 2005gl's position with Keck data, we looked at a Hubble image and saw something there that looked like a star, although we could not be sure. If it was a single star, its brightness (perhaps a million times that of the sun) suggested it was massive—100 times the sun's mass. Yet given prevailing opinion that such a heavyweight should not explode at all, most astronomers would have thought it more plausible that the dot of light in the Hubble image came from a cluster of smaller, fainter stars that together produced the brightness we saw. Our data could not rule out this possibility—yet.

Even though our result was inconclusive, I became increasingly interested in finding observational evidence speaking to the fate of the most massive stars. But science rarely follows a straight line from asking a question to finding an answer. I was thinking of stellar explosions of an entirely different kind—those called gamma-ray bursts—when a chance event in 2006 led to another surprising finding, which suggested not only that giant stars might go supernova but also that they could do so in a startling way.

ANOTHER STRANGE BLOWUP

THIS NEW CHAPTER in the story began with another night at the Keck observatory in 2006. This time, however, the gods seemed much less kind: the weather was terrible. I sat by the control computer and waited, as hours went by. Just as I was beginning to wonder whether my long trip back had been in vain, the clouds thinned out. The sky did not exactly clear up, but you could see some stars. I decided to observe the brightest supernova explosion visible at that time, an unusually luminous event called SN 2006gy, which then University of Texas at Austin graduate student Robert Quimby had discovered eight days earlier using a telescope less than one-twentieth the size of the giant Keck reflectors. I managed to observe for 15 minutes until the clouds thickened again, this time for good. It seemed like the night was a total loss.

But later a team led by my Caltech colleague Eran Ofek analyzed the data I had obtained, and SN 2006gy turned out to be the most luminous supernova explosion ever found to date. A parallel study led by Nathan Smith, then at the University of California, Berkeley, came to a similar conclusion. It made no sense. None of the types of supernovae we were aware of could generate so much light. SN 2006gy was in a galaxy that had not been imaged by Hubble before, so we also had no way of studying its progenitor star in detail. Judging from the violence of its explosion, though, the star probably weighed at least 100 solar masses.

We thought of several possible explanations for the luminosity, two of which seemed the least implausible. The first was that the extremely bright light was heat radiation from a shock wave that formed as the supernova's explosive debris caught up with the slower stellar wind that the star itself had emitted before exploding and swept that stellar wind away. The second option we considered was radioactivity. Supernovae synthesize new elements, largely in the form of radioactive isotopes that later decay into other, more stable ones. Perhaps this giant explosion synthesized a huge amount of radioactive material, whose slow decay injected energy into an expanding cloud of stellar debris and made the cloud glow in fluorescent light. But what could produce enough radioactive material to explain such outrageous luminosity?

That last question grabbed our interest. To try to answer it, we began to review past theoretical work. We stumbled on old, dusty theoretical papers from the late 1960s by three young astrophysicists—Gideon Rakavy, Giora Shaviv and Zalman Barkat. They had proposed a new way that a star could blow up.

Stars shine because their cores are dense and hot enough that hydrogen atoms fuse, turning into helium and heavier elements and releasing energy. Density and temperature, by and large, control the physics of the core of a massive star and the star's evolution. In general, as time progresses, the core gets denser and hotter. The core then crosses successive thresholds toward the fusion of increasingly heavy elements—first helium to carbon, then carbon to oxygen, and so on. Each stage between thresholds may last thousands to billions of years, depending on how fast the star's nuclear burning affects its core temperature and pressure.

EXTREME PHYSICS

IN BRIEF

In recent years several supernovae have proved to be more powerful and long-lasting than any seen before.
Archival images showed that the stars that gave rise to some supernovae were about 100 times as massive as the sun: according to accepted theory, stars this big were not supposed to explode.
Some supernovae may have been thermonuclear explosions triggered by the creation of electron-positron pairs of particles of matter and antimatter.
The first generation of stars in the universe, which created the materials that later formed planets, may have exploded through a similar mechanism.

© 2013 Scientific American

Rakavy and company calculated what would happen when a very massive star, perhaps hundreds of times more massive than the sun, reaches the stage at which its core is mostly oxygen. In lesser stars, we know what is next: the star contracts, and its core heats up until conditions allow the nuclear fusion of oxygen into silicon. But in a hypergiant, the theory said, the core would contract under gravity and heat up without becoming very dense. So instead of oxygen fusion, something else would happen: physicists call it pair production.

In matter that is hot enough, energetic particles such as nuclei and electrons emit very powerful light—photons so energetic that they are in the gamma-ray spectrum. Because of Albert Einstein's famous equation relating mass and energy, $E = mc^2$, two very energetic photons can, if they collide, spontaneously convert into pairs of other particles; specifically, they can transform into a pair that consists of an electron and its antiparticle, the positron. Most of the energy of the photons thus gets locked up in the form of matter. Consequently, electrons and positrons exert much lower pressure than the photons they originated from: they are deadweight. If the core of a very massive star reaches these conditions, its pressure suddenly falls, almost as if the star had a release valve. Before, pressure was what kept the star from collapsing under its own weight; now the core becomes unstable and begins to rapidly contract.

As density shoots up, it ignites the fusion of oxygen. Because the threshold to fusing oxygen is crossed in a collapsing core rather than a stable one, the ignition is explosive: fusion releases nuclear energy that heats the material further, which in turn speeds up the fusion, in a "runaway" reaction. The star can burn so much oxygen in such a short time—mere minutes—that the energy it releases is larger than the star's entire gravitational energy. Thus, whereas typical supernovae leave behind charred remains, in this type of explosion the object completely obliterates itself. All that is left is a fast-expanding cloud, much of it made of elements that were synthesized in the fury of the deflagration.

The theorists predicted that this type of event—called a pair-instability supernova because it destabilizes the star through the production of electron-positron pairs—would form a huge amount of nickel 56 in addition to other relatively heavy elements. Nickel 56 is an isotope with a tightly bound nucleus that nonetheless is radioactive, ultimately producing nonradioactive iron. If this scenario occurred in the precursor of SN 2006gy, we thought, the decay of nickel 56 might explain the supernova's intense luminosity.

Although the three astrophysicists' theory was correct, for decades common wisdom was that their hypothetical process would not actually take place in nature. Theorists who work on the formation and evolution of stellar bodies thought that such massive stars should not form at all, at least not in the present-day universe. And even if they did form, they would drive such strong stellar winds that they would rapidly lose most of their mass, leaving them unable to form cores massive enough to reach pair instability. The situation was different less than a billion years after the big bang. Then, the first stars might have been massive enough to explode as pair-instability supernovae. Perhaps.

Meanwhile the new record-smashing supernova, SN 2006gy, became a hit among astronomers, spurring more observational and theoretical studies. Ironically, even though SN 2006gy prompted us and others in the supernova community to reconsider the pair-instability model, this particular event did not, in the end, seem to have the right signature for nickel radioactivity—namely, a specific way the light dimmed with time. In a pair-instability explosion, most of the light should come not from the blast itself but from nickel 56 and the other radioactive isotopes it forges. Radioactivity is a well-studied process in which decay proceeds at a predictable, gradual rate. But SN 2006gy, after being bright for many months, quite suddenly disappeared, too quickly to have been powered by radioactivity. It was likely not a pair-instability supernova after all, and the other option we had considered—that the event's unusual brightness originated from a shock wave—became the accepted explanation. Still, the near miss had put me on the alert for signs of pair-instability events.

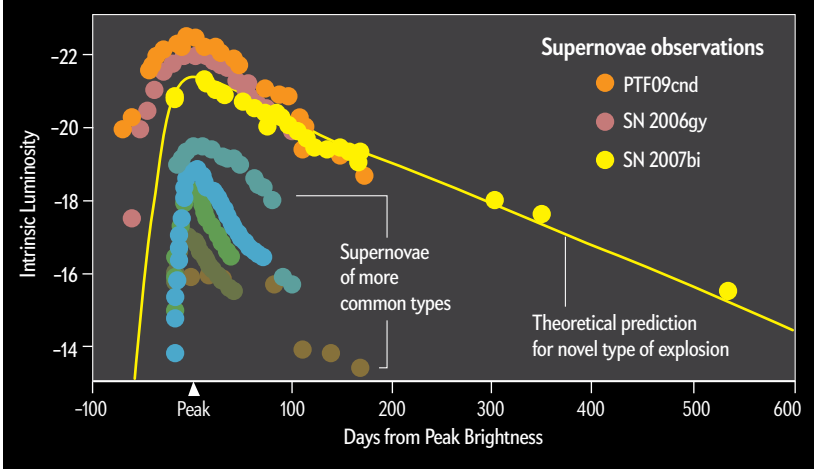
THE REAL THING?

A FEW MONTHS AFTER our lucky break with the Hawaiian clouds, I went to Colorado on vacation. Soon, however, I was interrupted by an e-mail from Peter Nugent of

SUPERNOVA OBSERVATIONS

The Brightest of the Bright

Supernova explosions studied by the author and his collaborators in the past few years have turned out to be the most energetic ever observed. One event, which began in 2006, reached record brightness (*pink*), beaten by another in 2009 (*orange*). But those died off relatively fast. Another one, from 2007, did not peak quite as high but released the most energy overall (*yellow*). It was the first example of a new type of explosion believed to occur in very massive stars [*see box on next two pages*].



SOURCE: AVISHAY GAL-YAM

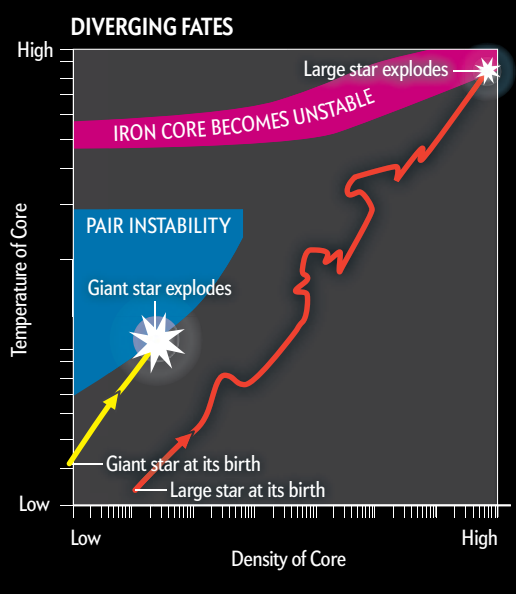
Graphics by Jen Christiansen

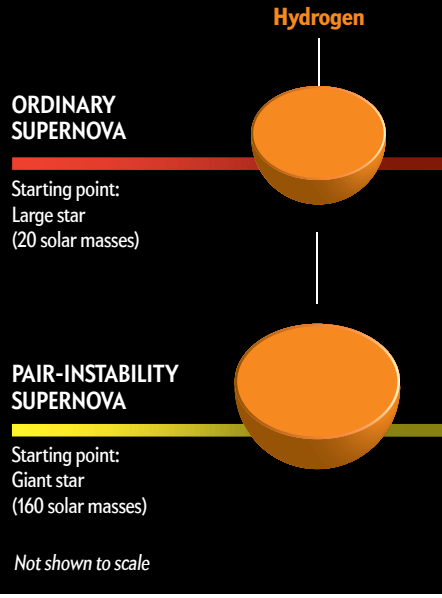
© 2013 Scientific American

SUPERNOVAE COMPARED

How Large Stars Die

Stars forge new elements by nuclear fusion, which is what makes them shine. As a star ages, its core gets hotter and denser (*graph*) and produces heavier and heavier elements, which tend to form onionlike layers (*diagram*). A relatively heavy star, such as one of 20 solar masses (*red in graph and diagram*), eventually becomes dense enough that it collapses, spewing out large amounts of energy and much of its mass. But a very heavy star, say, 160 solar masses (*yellow*), annihilates itself sooner in a recently discovered, even mightier type of blast.





EXTREME PHYSICS

Lawrence Berkeley National Laboratory. Nugent and I had just started a "practice run" for a big supernova search we had been planning. Now he sent me a supernova with a weird spectrum. I had never seen its like before.

Because atoms of each element in nature absorb and emit light at particular wavelengths, the spectrum of an astronomical source provides information about the chemical composition of the material emitting the light. The spectrum of Nugent's object—SN 2007bi—suggested that the elements that composed it were present in unusual proportions and that it was extremely hot.

After I got back to Caltech, I continued to track the evolution of this event. It emitted about 10 times more light than the typical supernova. And the amount of light declined very slowly: this source refused to fade away, as days turned into weeks and weeks into months. I became more and more convinced that this was finally an example of a pair-instability supernova. It took more than a year before it finally disappeared. But I needed more data to be sure of my interpretation.

During 2007 and 2008 several collaborators and I continued to observe SN 2007bi using telescopes at Caltech's Palomar Observatory. As the light from this explosion finally grew fainter, about a year after we discovered it, I asked my Caltech colleagues Richard Ellis and Kulkarni to observe it with the large telescopes at Keck—promising in my e-mails that this was "the real deal."

In the meantime, I moved to Israel with my family and took up my current job at the Weizmann Institute of Science in Rehovot. In August 2008 Kulkarni and his graduate student Mansi Kasliwal sent me the latest spectrum for SN 2007bi. When I did a first, rough analysis, I could not believe what I saw. I analyzed the spectrum over and over, but the answer was the same: this explosion synthesized a staggering amount of nickel 56: between five and seven times the entire mass of our sun. It was 10 times more than we or anyone else had ever seen before—and just what you expect from a pair-instability supernova explosion. That night I paced back and forth in my apartment, thinking about this finding and its implications. When my wife gave me a strange look and asked what was going on, I said, "I think we've made a great discovery."

In late 2008 I traveled to Garching, Germany, to work with Paolo Mazzali at the Max Planck Institute for Astrophysics. Mazzali is a world expert in quantitative analysis of supernova spectra, so he could test the results of my rough analysis. He also had additional useful data he had obtained with another large instrument, the European Southern Observatory's Very Large Telescope in Chile. We sat together in his office as Mazzali ran his code. Yes! The results were consistent with my previous analyses: many solar masses of nickel 56 and a relative abundance of elements matching the predictions of pair-instability models.

DOUBLE TAKE

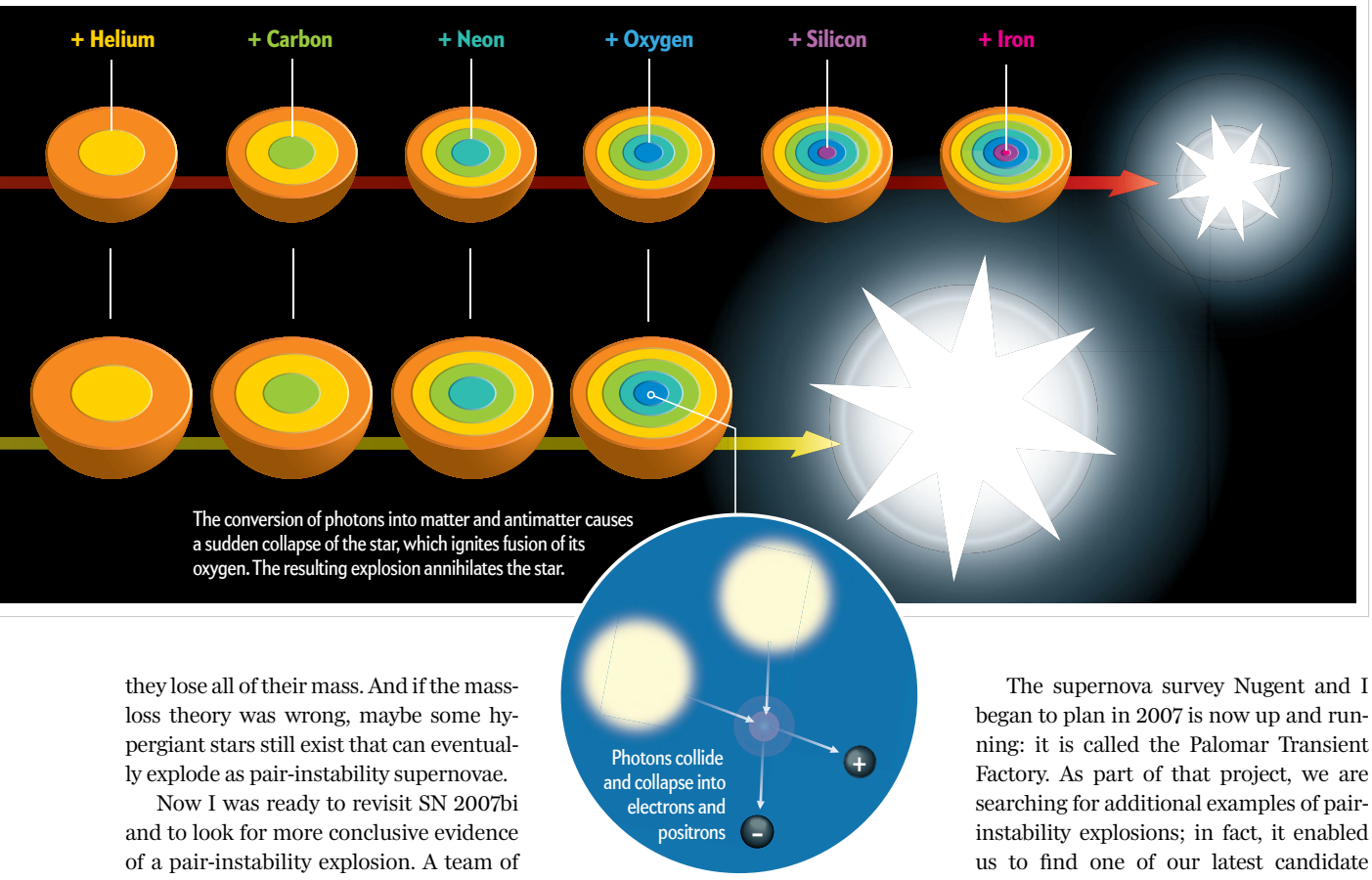
ALTHOUGH I WAS PRETTY CONFIDENT that we had identified a pair-instability supernova, when I returned to Israel I set the data aside for a few months while I was busy on another project involving the supernova that had set me on this journey in the first place: SN 2005gl. When Fox, Leonard and I had found its putative progenitor star in late 2005, we could not be positive that it was indeed a single entity rather than a cluster of stars. Now, three years later, the supernova had disappeared, and I realized we could do a simple test: if our candidate was not the star that had blown up, it would still be there. Leonard and I returned to the Hubble to check.

By the end of 2008 we were finally sure: the star had disappeared. The progenitor of SN 2005gl was indeed very luminous and probably quite massive—a twin of Eta Carinae, one of the heftiest blue giants in our own galaxy.

Thus, the prevailing theory of hypergiant stars—that they lose most of their mass before they explode—was wrong at least in this case. Some very luminous and massive stars do exist and explode before

SOURCE: MODIFIED FROM "THE MOST MASSIVE CORE-COLLAPSE SUPERNOVA PROGENITORS," BY R. WALDMAN, IN *ASTROPHYSICAL JOURNAL*, VOL. 685; OCTOBER 1, 2008 (*graph*)

© 2013 Scientific American


+ Helium
+ Carbon
+ Neon
+ Oxygen
+ Silicon
+ Iron
The conversion of photons into matter and antimatter causes a sudden collapse of the star, which ignites fusion of its oxygen. The resulting explosion annihilates the star.
Photons collide and collapse into electrons and positrons
+
–

they lose all of their mass. And if the mass-loss theory was wrong, maybe some hypergiant stars still exist that can eventually explode as pair-instability supernovae.

Now I was ready to revisit SN 2007bi and to look for more conclusive evidence of a pair-instability explosion. A team of collaborators and I tested it in every way we could think of. We analyzed its spectra in detail and how its light evolved in time. We compared old models of stellar explosion and new ones. Near the end of 2009 all the evidence converged into a single conclusion: the most logical, almost inescapable way to explain SN 2007bi was that it was a pair-instability supernova. After more than two years of study, it was finally time to start publishing our results.

We have now collected three more events that are strong candidates for pair-instability supernovae. Overall, they appear to be exceedingly rare—constituting only one out of 100,000 supernovae—and to require a star of at least 140 solar masses and perhaps as many as 200. But they are huge factories of the elements, and they produce the most energetic explosions known to science. They might even deserve the name "hypernovae."

Perhaps the most fascinating aspect of this new type of supernova is that it gives us a glimpse into the early universe. The very first stars to light up, some 100 million years after the big bang, would have measured upward of 100 solar masses and maybe as much as 1,000. Some of those behemoths probably exploded via a pair-instability mechanism. Thus, the distant cousins of some of today's supernovae may have been the first explosions to seed the universe with heavier elements, thereby shaping the stars and planets that followed them—including our sun and Earth.

Not only do our observations suggest a novel way for stars to blow up, they also mean that the modern universe, contrary to earlier views, probably is sprinkled with hypergiant stars. Growth to extraordinary sizes for primordial stars was possible only in an environment made almost exclusively of hydrogen and helium. "Pollution" with the products of nuclear fusion then put a choke hold on stellar accretion: in the presence of heavier elements, stars collapse faster and thus ignite sooner, blowing off any residual gas around them before they can grow too heavy. But clearly, the heavier elements are less of a brake on stellar growth than astrophysicists used to believe.

The supernova survey Nugent and I began to plan in 2007 is now up and running: it is called the Palomar Transient Factory. As part of that project, we are searching for additional examples of pair-instability explosions; in fact, it enabled us to find one of our latest candidate events, which looks very much like SN 2007bi. As data accumulate, our understanding of these explosions and how they contribute to making the heavy elements in the universe deepens. Future instruments, such as NASA's next-generation observatory, the James Webb Space Telescope, will probably be able to detect very distant pair-instability explosions. Perhaps one day they will reveal the explosive deaths of the first stars to have ever formed in our universe. SA

Avishay Gal-Yam graduated with a Ph.D. in astrophysics in 2004 at Tel Aviv University and has been a Hubble Postdoctoral Fellow at the California Institute of Technology. He is now an associate professor of physics at the Weizmann Institute of Science in Rehovot, Israel.

MORE TO EXPLORE

How to Blow Up a Star. Wolfgang Hillebrandt, Hans-Thomas Janka and Ewald Müller in *Scientific American*, Vol. 295, No. 4, pages 42–49; October 2006.

A Massive Hypergiant Star as the Progenitor of the Supernova SN 2005gl. A. Gal-Yam and D. C. Leonard in *Nature*, Vol. 458, pages 865–867; April 16, 2009.

Supernova 2007bi as a Pair-Instability Explosion. A. Gal-Yam et al. in *Nature*, Vol. 462, pages 624–627; December 3, 2009.

© 2013 Scientific American

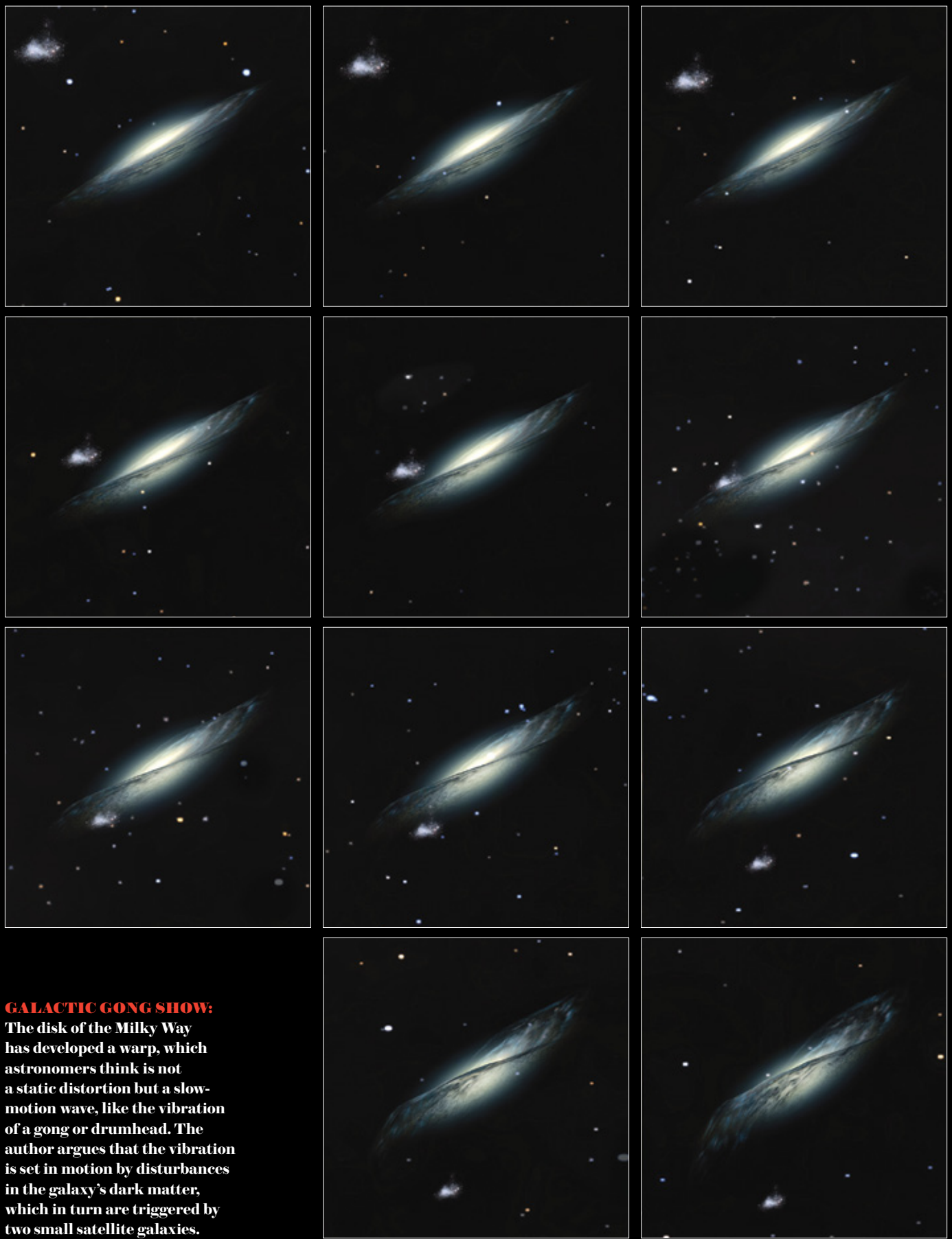
ASTRONOMY

THE DARK SIDE OF THE MILKY WAY

Dark matter is not just a puzzle. It is a solution

By Leo Blitz

© 2013 Scientific American



GALACTIC GONG SHOW: The disk of the Milky Way has developed a warp, which astronomers think is not a static distortion but a slow-motion wave, like the vibration of a gong or drumhead. The author argues that the vibration is set in motion by disturbances in the galaxy's dark matter, which in turn are triggered by two small satellite galaxies.

Illustrations by Don Dixon

© 2013 Scientific American

ALTHOUGH ASTRONOMERS ONLY SLOWLY CAME TO REALIZE dark matter's importance in the universe, for me personally it happened in an instant. In my first project as a postdoc at the University of California, Berkeley, in 1978, I measured the rotational velocities of star-forming giant molecular clouds in the outer part of the disk of our Milky Way galaxy. I worked out what was then the most accurate method to determine those velocities, and I sat down to plot out the results (by hand on graph paper) in the astronomy department lounge. Two other experts on the Milky Way, Frank Shu and Ivan King, happened by. They watched as I filled in the velocities of the outermost clouds, and the pattern we saw made it clear at once that the Milky Way was rife with dark matter, especially in its outermost parts. We sat and scratched our heads, imagining what the nature of the dark matter could be, and all the ideas we came up with turned out in short order to be wrong.

This study was one of many in the 1970s and 1980s that forced astronomers to conclude that dark matter—a mysterious substance that neither emits nor absorbs light and reveals itself solely by its gravitational influence—not only exists but is the dominant material constituent of the universe. Measurements with the Planck spacecraft confirm that dark matter accounts for 5.4 times as much mass as ordinary matter (protons, neutrons, electrons, and so on). What the stuff is remains as elusive as ever. It is a measure of our ignorance that the most conservative hypothesis proposes that dark matter consists of an exotic particle not yet detected in particle accelerators, predicted by theories of matter that have not yet been verified. The most radical hypothesis is that Newton's law of gravity and Einstein's general theory of relativity are wrong or, at the very least, require unpleasant modifications.

Whatever its nature, dark matter is already providing keys to unlock some persistent puzzles about how the Milky Way came to have certain of its features. For example, astronomers have known for more than 50 years that the outer parts of

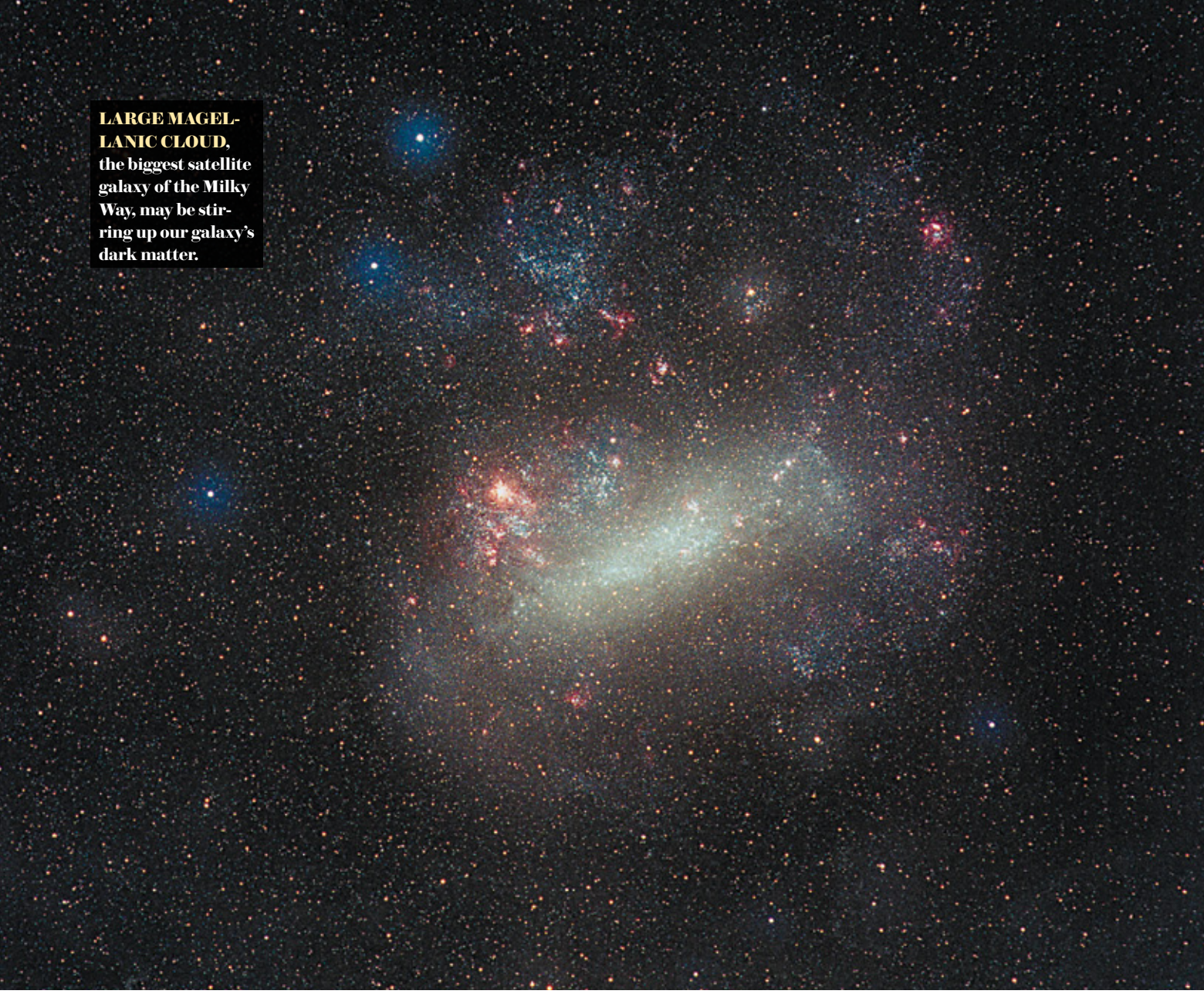
IN BRIEF

Dark matter is one of the great scientific riddles of our time, but once astronomers accepted the existence of dark matter, the answers to many other cosmic mysteries began to fall into place.

Whatever this unknown material may be, it seems to explain why the disk of our Milky Way galaxy has such a pronounced warp at its outer rim. Orbiting satellite galaxies naturally tend to distort the galaxy, but their gravitational effect would be too weak without the amplification that dark matter provides.

Another question dark matter answers is why the Milky Way appears to have fewer such satellite galaxies than models predict it should. It turns out that the satellites are probably out there, though composed almost entirely of dark matter, making them hard to detect.

© 2013 Scientific American



LARGE MAGELLANIC CLOUD, the biggest satellite galaxy of the Milky Way, may be stirring up our galaxy's dark matter.

COURTESY OF ECKHARD SLAWIK

the galaxy are warped like a vinyl phonograph record left on a heater. They could not make a viable model for the warp—until they considered the effects of dark matter. Similarly, computer simulations of galactic formation based on the assumed properties of dark matter predicted that our galaxy should be surrounded by hundreds or even thousands of small satellite galaxies. Yet observers saw only about two dozen. The discrepancy led people to question whether dark matter had the properties they thought it did. But in recent years several groups of astronomers have discovered troves of dwarf satellites, narrowing the disparity. These newly located satellites are not only helping to resolve a long-standing mystery of galactic structure; they may also be teaching us something about the total cosmic inventory of matter.

FACTORING IN THE WARP

A FIRST STEP to understanding what dark matter tells us about the Milky Way is to get a general picture of how the galaxy is organized. Ordinary matter—the stars and gas—resides in four major structures: a thin disk (which includes the pinwheel-like spiral pattern and the location of the sun), a dense nucleus (which also harbors a supermassive black hole), an elongated bulge known as the bar, and a spheroidal "halo" of old stars and clusters that envelops the rest of the galaxy. Dark matter has a very different arrangement. Although we cannot see it, we infer where it is from the rotational velocities of stars and gas; its gravitational effects on visible material suggest it is approximately spherically distributed and extends far beyond the stellar halo, with a density that is highest at the center and falls off approximately as the square of the distance from the center. Such a distribution would be the natural result of what astronomers call hierarchical merging: the proposition that in the early universe, smaller galaxies accreted to build larger ones, including the Milky Way.

For years astronomers could get no further than this basic picture of dark matter as a giant, undifferentiated ball of unidentified material. In the past several years, however, we have managed to glean more details, and dark matter has proved rather more interesting than

© 2013 Scientific American

we had suspected. Various lines of evidence suggest that this material is not smoothly distributed but has some large-scale lumpiness to it.

Such unevenness would explain the existence and size of the galactic warp. When astronomers say that the galaxy is warped, we are referring to a specific distortion in the outskirts of the disk. At distances beyond about 50,000 light-years from the center, the disk consists almost entirely of atomic hydrogen gas, with only a few stars. Mapped by radio telescopes, the gas does not lie in the plane of the galaxy; the farther out you go, the more it deviates. By a distance of about 75,000 light-years, the disk has bent about 7,500 light-years out of the plane [*see box at right*].

Evidently, as the gas within the disk revolves around the galactic center, it also oscillates up and down, in and out of the plane. These oscillations occur over hundreds of millions of years, and we catch them at one moment in their cycle. In essence, the gas disk acts as a kind of giant gong vibrating in slow motion. Like a gong, it can vibrate at multiple frequencies, each corresponding to a certain shape of the surface. In 2005 my colleagues and I showed that the observed warp is the sum of three such frequencies. (The lowest is 64 octaves below middle C.) The overall effect is asymmetric: gas on one side of the galaxy is much farther from the plane than gas on the other side.

The radio astronomers who first noticed the warping in the 1950s thought it might result from gravitational forces exerted by the Magellanic Clouds, the most massive galaxies in orbit around the Milky Way. Because these satellite galaxies are orbiting out of the plane of the Milky Way, their gravity tends to distort the disk. Detailed calculations, however, showed that these forces are too weak to explain the effect because the Magellanic Clouds are puny in comparison to the Milky Way. For decades the reason for the pronounced warp remained an unsolved problem.

DARK HAMMER

THE RECOGNITION that the Milky Way contains dark matter, together with new estimates of the mass of the Magellanic Clouds (which showed them to be more massive than thought), raised a new possibility. If the gas disk acts as a giant gong, the orbit of the Magellanic Clouds through the dark matter halo can act as a hammer ringing the gong, sounding its natural notes or resonant frequencies, albeit not directly. The clouds create a wake in the dark matter, just as a boat forms a wake as it plows through the water. In this way, the clouds create some unevenness in the distribution of dark matter. That, in turn, acts as the hammer to cause a ringing in the low-mass, outer parts of the disk. The upshot is that even though the Magellanic Clouds are puny, dark matter greatly amplifies their effects.

Martin D. Weinberg of the University of Massachusetts Amherst put forward this general idea in 1998. He and I later applied it to observations of the Milky Way and found we could reproduce the three vibration patterns of the gas disk. If the theory is correct, the warp is an active feature of the Milky Way with a shape that continually changes as the Magellanic Clouds move through their orbits. The shape of the galaxy is not fixed but ever shifting. [*Editors' note: A video of this process is available at www.ScientificAmerican.com/oct2011/blitz.*]

The warp is not the only asymmetry in the shape of the Milky Way. One of the most striking is the lopsided thickness of the outer gas disk, also discovered using radio telescopes. If one drew a line from the sun to the center of the Milky Way and extended it outward, one would find that the thickness of the gas layer on one side of this line is, on average, about twice that on the other side. This large asymmetry is dynamically unstable and, left to its own devices, would tend to right itself; its persistence requires some sort of mechanism to maintain it. For 30 years astronomers knew about the problem but swept it under the rug. They revisited it only recently when a much improved new survey of the Milky Way's atomic hydrogen, coupled with a better understanding of the noncircular motions of the gas, made the asymmetry impossible to ignore any longer.

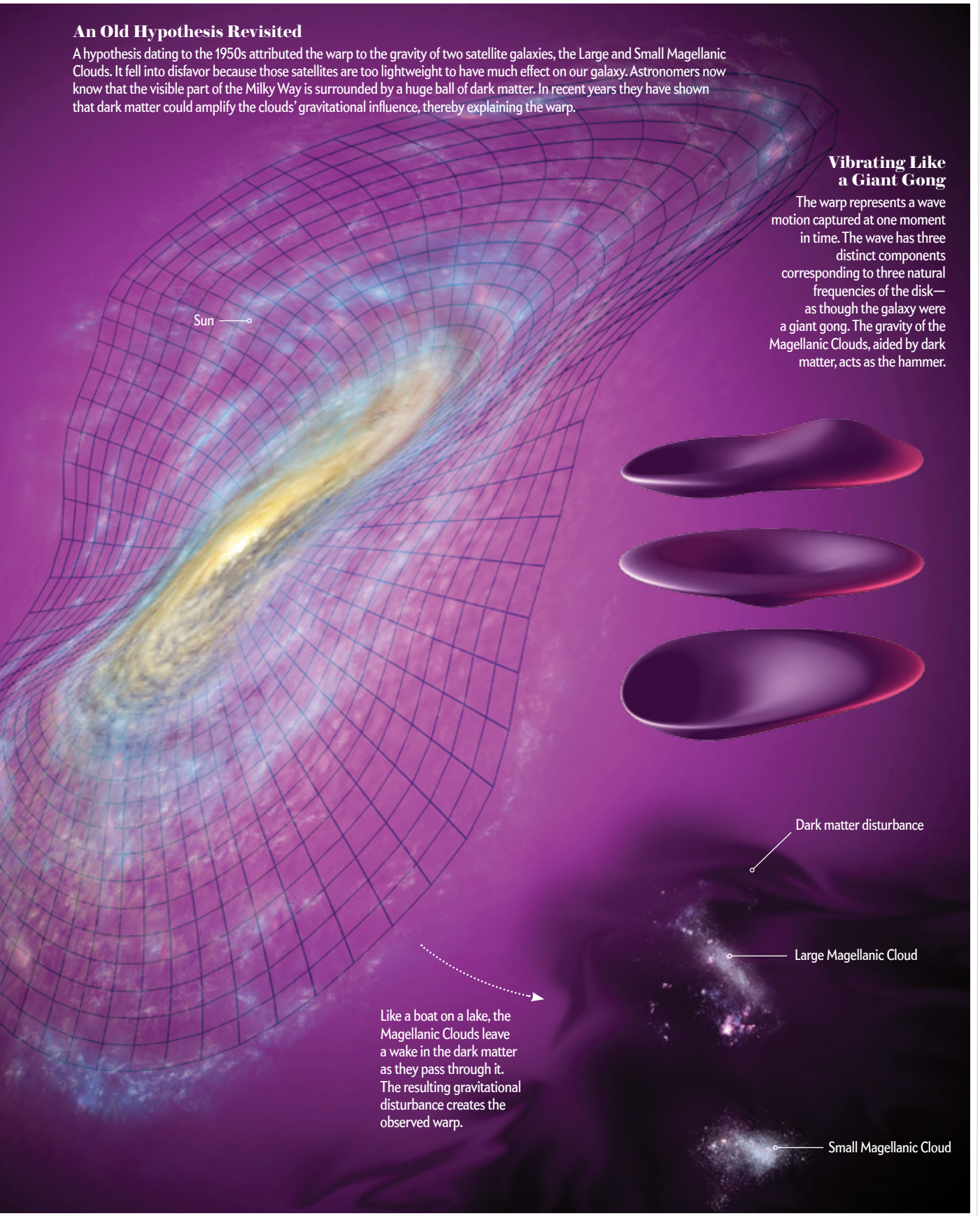
The two leading explanations both involve dark matter. Either the Milky Way is spherical but not concentric with its dark matter halo, as recently suggested by Michael Kuhlen of U.C. Berkeley and his colleagues, or the dark matter halo is itself somewhat asymmetric, as suggest-

SOLVED: MYSTERY #1

Galactic Warp Factor
The disk of our Milky Way galaxy, which contains most of its stars and gas, has roughly the proportions of a vinyl LP record or compact disk. The warp in the galaxy looks like what happens when you mistreat an LP or CD.
Classic view of Milky Way
Some warp explained by Magellanic Clouds
Magellanic Clouds
Dark matter halo (purple haze)
Vertical scale exaggerated

SOURCE: GYULA JÓZSA (*warp shape insets*)

© 2013 Scientific American


An Old Hypothesis Revisited
A hypothesis dating to the 1950s attributed the warp to the gravity of two satellite galaxies, the Large and Small Magellanic Clouds. It fell into disfavor because those satellites are too lightweight to have much effect on our galaxy. Astronomers now know that the visible part of the Milky Way is surrounded by a huge ball of dark matter. In recent years they have shown that dark matter could amplify the clouds' gravitational influence, thereby explaining the warp.
Vibrating Like a Giant Gong
The warp represents a wave motion captured at one moment in time. The wave has three distinct components corresponding to three natural frequencies of the disk—as though the galaxy were a giant gong. The gravity of the Magellanic Clouds, aided by dark matter, acts as the hammer.
Sun
Dark matter disturbance
Large Magellanic Cloud
Like a boat on a lake, the Magellanic Clouds leave a wake in the dark matter as they pass through it. The resulting gravitational disturbance creates the observed warp.
Small Magellanic Cloud

© 2013 Scientific American

ed by Kanak Saha of the Max Planck Institute for Extraterrestrial Physics in Garching and several collaborators. Both explanations call into question astronomers' old view that the Milky Way and the halo formed together from the condensation of a single gargantuan cloud of material; if it had, the ordinary matter and the dark matter should be centered on the same point. Therefore, the asymmetry is further evidence that the galaxy formed from the merging of smaller units or grew by continued merging or accretion of intergalactic gas—processes that need not be symmetric. The center of the galaxy could be offset from the center of the dark matter because gas, stars and dark matter behave differently.

A way to cross-check this idea is to study the long, thin streams of stars that stretch through the outer reaches of the galaxy. These formations are the elongated remains of former satellite galaxies. The most common kind of galaxy to be found in orbit about the Milky Way system is known as a dwarf spheroidal because of its roundish shape and small mass of stars—typically only about one ten-thousandth that of the Milky Way. Over time its orbit decays, and the satellite becomes subject to the tidal forces of the Milky Way. These forces are the same as those produced by the moon on Earth, stretching out the mass of water on Earth as our planet rotates, producing the twice-daily ocean tides. The dwarf galaxy gets stretched out and can be reduced to a thin ribbon.

Because the stars in these streams orbit the galaxy at large distances, where the gravitational effects of dark matter are large, the shapes of the streams probe the shape of the halo. If the halo is not perfectly spherical but somewhat flattened, it will exert a torque on the orbits of stars in the stream and cause a marked deviation from a great circle. As it happens, the streams are observed to be very thin, and their orbits around the galaxy are nearly great circles. Computer simulations done by Rodrigo Ibata of the Strasbourg Astronomical Observatory in France and his colleagues thus suggest that the dark matter distribution is close to spherical, although it might nonetheless be as lopsided as suggested by Saha and his colleagues.

GALAXIES GONE MISSING

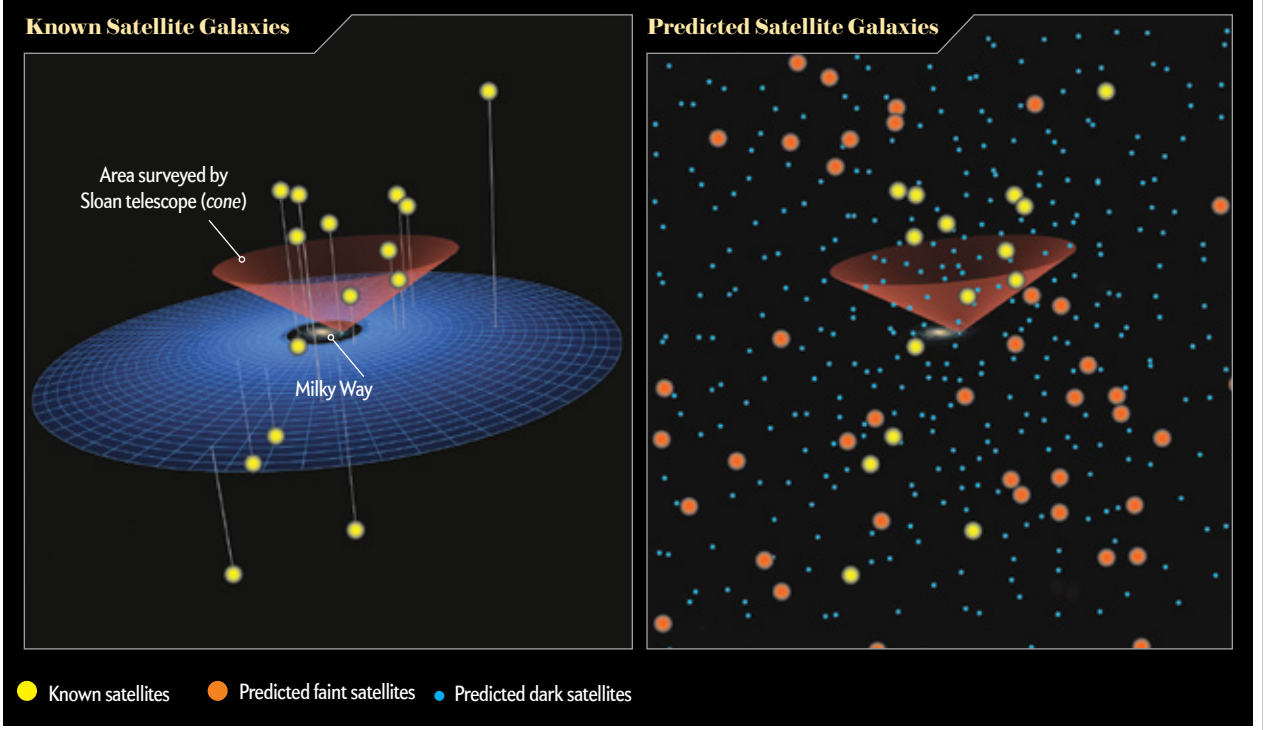
IF THE DESTRUCTION of dwarf galaxies raises questions, so does their formation. In our current models, galaxies begin as agglomerations of dark matter, which then accrete gas and stars to form their visible parts. The process yields not only large galaxies such as ours but also numerous dwarfs. The models get the properties of these dwarfs about right but predict far more of them than observers see. Does the fault lie in the models or with the observations?

Part of the answer comes from new analyses of the Sloan Digital Sky Survey, a systematic scan of about a quarter of the sky. The survey has turned up about

ALMOST SOLVED: MYSTERY#2

Lost Sheep of the Galactic Family

Theories predict our Milky Way should be orbited by hundreds of satellite galaxies. Astronomers have long worried they could find only two dozen or so, but new searches using the Sloan Digital Sky Survey have closed the gap by spotting previously unseen satellites. They are composed almost entirely of dark matter. (The positions of the predicted satellites are schematic, reflecting their overall distribution.)



SOURCE: JAMES BULLOCK, *University of California, Irvine (known satellites)*

© 2013 Scientific American

16 new, extremely dim galaxies in orbit around the Milky Way. Their discovery is astonishing. The sky has been so completely surveyed for so long that it is difficult to imagine how galaxies on our cosmic doorstep could have lain undiscovered all this time. These galaxies, known as ultrafaint dwarfs, in some cases contain only a few hundred stars. They are so feeble and diffuse that they do not show up on ordinary images of the sky; it requires special data-handling techniques to identify them.

Had the Sloan survey covered the entire sky when the ultrafaint galaxies were discovered, it might have discovered another 35 or so more. Still, that would not account for all the "missing" dwarfs. So astronomers have sought other possibilities. Perhaps more such galaxies are out there, too far away for existing telescopes to detect. The Sloan survey can find ultrafaint dwarfs out to a distance of about 150,000 light-years. Erik Tollerud and his collaborators at the University of California, Irvine, predict that as many as 500 undiscovered galaxies orbit the Milky Way at distances up to around one million light-years from the center. Astronomers should be able to find them with a new optical telescope called the Large Synoptic Survey Telescope, which will have eight times the collecting area of the Sloan telescope. Construction began on the observatory in March 2011.

Another hypothesis is that the Milky Way is orbited by galaxies even dimmer than the faintest ultrafaint dwarfs—so dim, perhaps, that they contain no stars at all. They are almost pure dark matter. Whether such galaxies could ever be seen depends on whether they contain gas in addition to the dark matter. Such gas might be sufficiently diffuse that it cools only very slowly, too slowly to have formed stars. Radio telescopes surveying large patches of the sky might nonetheless detect the gas.

If these galaxies lack gas, however, they would reveal their presence only indirectly, by their gravitational effects on ordinary matter. If one of these dark galaxies hurtled through the disk of the Milky Way or some other galaxy, it might leave a splash like that of a pebble thrown into a quiet lake—observable as perturbations to the distribution or velocities of stars and gas. Unfortunately, this splash would be very small, and astronomers would have to convince themselves that it could not be made in any other way—a daunting task. All spiral galaxies show disturbances throughout their atomic hydrogen disks akin to waves in a rough sea.

If the dark galaxy is massive enough, a method devised by Sukanya Chakrabarti, now at Florida Atlantic University, and several collaborators, including myself, may provide the tools to discern its passage. We recently showed that the largest disturbances in the outskirts of galaxies are often tidal imprints left by passing galaxies, which can be differentiated from other perturbations. By analyzing the disturbances, we can infer both the mass and current location of the intrusive galaxy.

This technique can discern galaxies as small as one-thousandth the mass of the primary galaxy. Applying this method to the Milky Way, our team inferred that an undiscovered possibly dark galaxy lurks in the plane of the Milky Way, about 300,000 light-years from the galactic center. Astronomers have begun hunting for this galaxy in near-infrared light using data collected by the Spitzer Space Telescope.

TOO LITTLE LIGHT

QUITE APART from the challenge of finding them, ultrafaint and dark galaxies in the Milky Way's vicinity pose a deeper problem for astronomers in regard to the relative amounts of material they contain. Astronomers commonly measure the amount of matter in a galaxy in terms of its mass-to-light ratio: the mass of material divided by the total amount of light it gives off. Typically we give the ratio in solar units; the sun, by definition, has a mass-to-light ratio of 1. In our galaxy, the average star is somewhat less massive and much dimmer than the sun, so the overall mass-to-light ratio of luminous matter is closer to 3. Including dark matter, the total mass-to-light ratio of the Milky Way jumps to about 30.

Josh Simon, now at the Carnegie Institution of Washington, and Marla Geha, now at Yale University, measured the velocities of the stars in eight ultrafaint dwarfs to obtain the mass and luminosity of these galaxies. The mass-to-light ratios in some cases exceed 1,000—by far the highest of any structure in the known universe. In the universe as a whole, the ratio of dark to ordinary matter is almost exactly 5.4. Why is the mass-to-light ratio of the Milky Way system so much higher and the ultrafaint galaxies even more so?

The answer could lie in the numerator or denominator of the ratio: galaxies with mass-to-light ratios that are higher than the universal average either have more mass than expected or produce less light. Astronomers speculate that the denominator is to blame. A huge amount of ordinary matter does not radiate brightly enough to see, either because it has never been able to settle into galaxies and coalesce into stars or because it did settle into galaxies but was then expelled back out into intergalactic space, where it resides in an ionized form that is undetectable by present-day telescopes.

Lower-mass galaxies, having weaker gravity, lose more of their gas, so their light output is disproportionately reduced. What a curious irony that the problems raised by one kind of unseen matter (dark matter) should give rise to yet another set (ordinary but undetected matter).

The puzzle of dark matter, which lay dormant for so many years, is now one of the most vibrant research areas in both physics and astronomy. Physicists are hoping to identify and detect the particle that composes dark matter, and astronomers are looking for more clues about how the stuff behaves. But puzzle or no, the existence of dark matter has provided the answer to a large range of astronomical phenomena. SA

Leo Blitz wanted to be an astronomer since early elementary school, when his favorite TV show was *Watch Mr. Wizard*. He is now a professor at the University of California, Berkeley, and is former director of the Radio Astronomy Laboratory.

MORE TO EXPLORE

A Magellanic Origin for the Warp of the Galaxy. Martin D. Weinberg and Leo Blitz in *Astrophysical Journal Letters*, Vol. 641, No. 1, pages L33–L36; April 10, 2006. http://arxiv.org/abs/astro-ph/0601694

The Vertical Structure of the Outer Milky Way H_I Disk. E. S. Levine, Leo Blitz and Carl Heiles in *Astrophysical Journal*, Vol. 643, No. 2, pages 881–896; June 1, 2006. http://arxiv.org/abs/astro-ph/0601697

Finding Dark Galaxies from Their Tidal Imprints. Sukanya Chakrabarti, Frank Bigiel, Philip Chang and Leo Blitz in *Astrophysical Journal*, Vol. 743, No. 1; December 2010. http://m.iopscience.iop.org/0004-637X/743/1/35

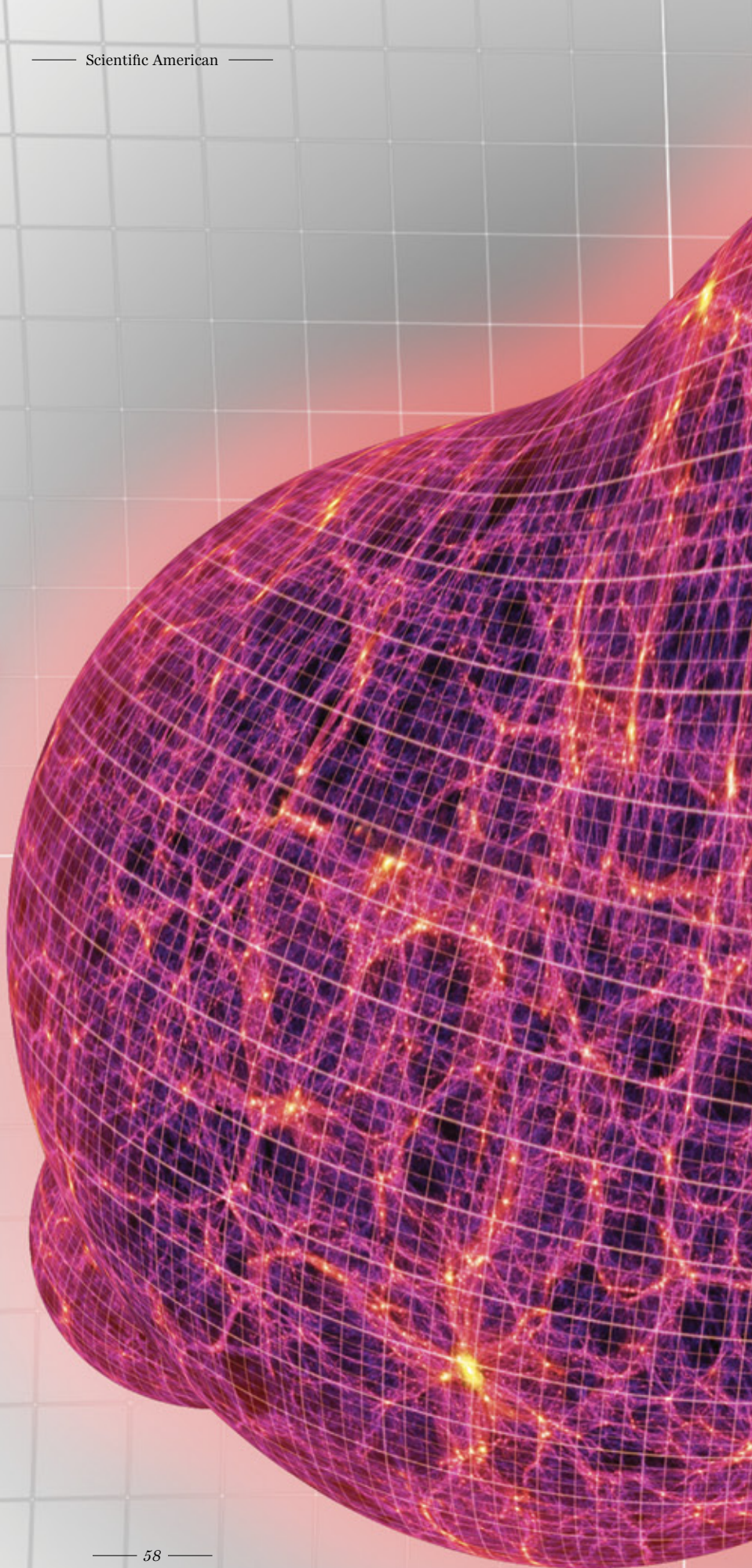
© 2013 Scientific American

COSMIC STRUCTURE

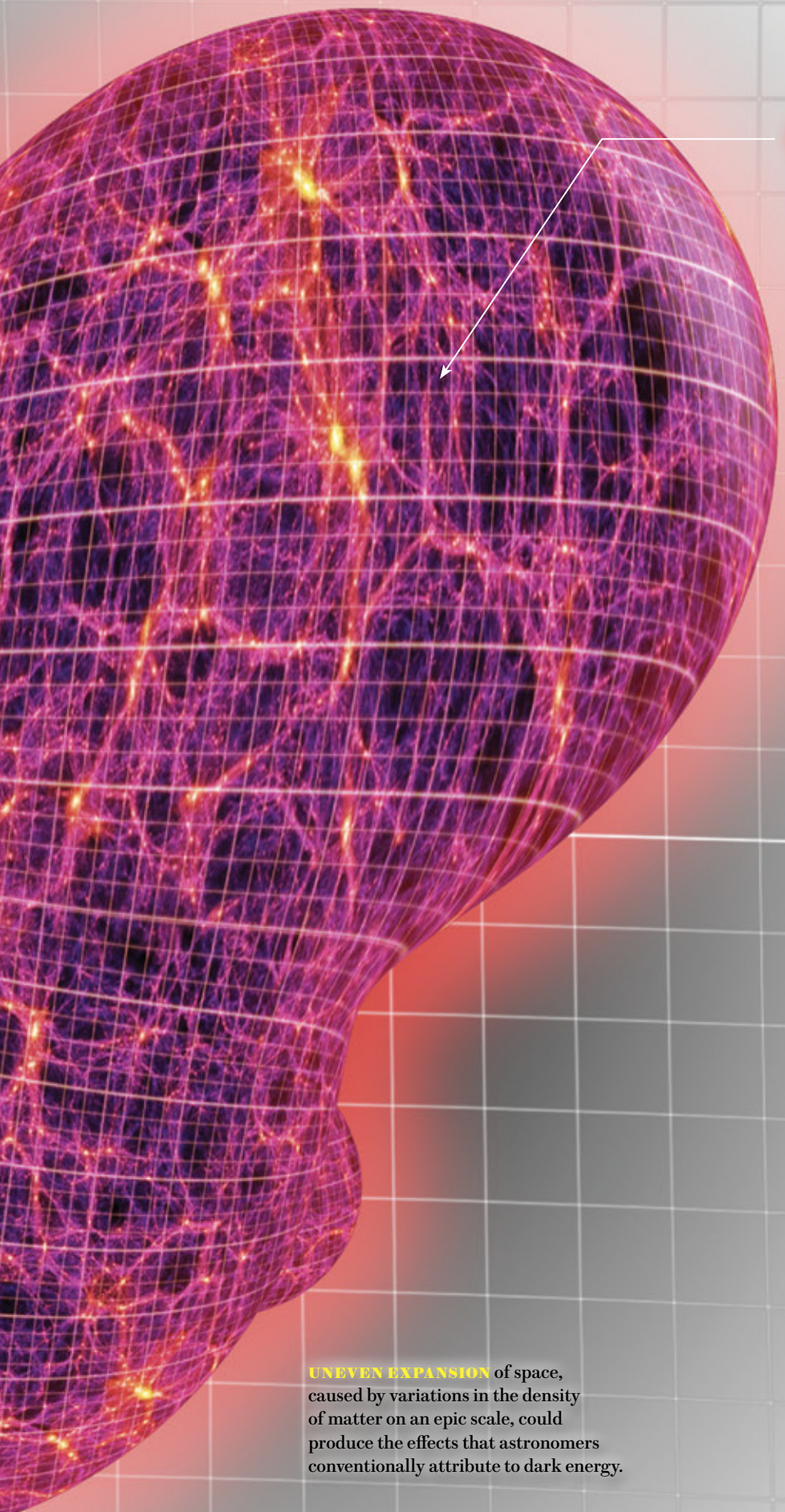
DOES DARK ENERGY REALLY EXIST?

Maybe not. The observations that led astronomers to deduce its existence could have another explanation: our galaxy lies at the center of a giant cosmic void

By Timothy Clifton and Pedro G. Ferreira



© 2013 Scientific American



UNEVEN EXPANSION of space, caused by variations in the density of matter on an epic scale, could produce the effects that astronomers conventionally attribute to dark energy.

Illustrations by Don Dixon

IN SCIENCE, OFTEN THE GRANDEST revolutions are triggered by the smallest discrepancies. In the 16th century, based on what struck many of his contemporaries as the esoteric minutiae of celestial motions, Copernicus suggested that Earth was not, in fact, at the center of the universe. In our own era, another revolution began to unfold 15 years ago with the discovery of the accelerating universe. A tiny deviation in the brightness of exploding stars led astronomers to conclude that they had no idea what 70 percent of the cosmos consists of. All they could tell was that space is filled with a substance unlike any other—one that pushes along the expansion of the universe rather than holding it back. This substance became known as dark energy.

The existence of dark energy is still so puzzling that some cosmologists are revisiting the fundamental postulates that led them to this deduction in the first place. One of these is the product of that earlier revolution: the Copernican principle, that Earth is not in a central or otherwise special position in the universe. If we discard this basic principle, a surprisingly different picture of what could account for the observations emerges.

Most of us are very familiar with the idea that our planet is nothing more than a tiny speck orbiting a typical star, somewhere near the edge of an otherwise unnoteworthy galaxy. In the midst of a universe populated by billions of galaxies that stretch out to our cosmic horizon, we are led to believe that there is nothing special or unique about our location. But what is the evidence for this cosmic humility? And how would we be able to tell if we *were* in a special place? Astronomers typically gloss over these questions, assuming our own typicality sufficiently

© 2013 Scientific American

obvious to warrant no further discussion. To entertain the notion that we may, in fact, have a special location in the universe is, for many, unthinkable. Nevertheless, that is exactly what some small groups of physicists around the world have recently been considering.

Ironically, assuming ourselves to be insignificant has granted cosmologists great explanatory power. It has allowed us to extrapolate from what we see in our own cosmic neighborhood to the universe at large. Huge efforts have been made in constructing state-of-the-art models of the universe based on the cosmological principle—a generalization of the Copernican principle that states that at any moment in time all points and directions in space look the same. Combined with our modern understanding of space, time and matter, the cosmological principle implies that space is expanding, that the universe is getting cooler and that it is populated by relics from its hot beginning—predictions that are all borne out by observations.

Astronomers find, for example, that the light from distant galaxies is redder than that of nearby galaxies. This phenomenon, known as redshift, is neatly explained as a stretching of light waves by the expansion of space. Also, microwave detectors reveal an almost perfectly smooth curtain of radiation emanating from very early times: the cosmic microwave background, a relic of the primordial fireball. It is fair to say that these successes are in part a result of our own humility—the less we assume about our own significance, the more we can say about the universe.

DARKNESS CLOSES IN

SO WHY ROCK THE BOAT? If the cosmological principle is so successful, why should we question it? The trouble is that recent astronomical observations have been producing some very strange results. Astronomers have found that for a given redshift, distant supernova explosions look dimmer than expected. Redshift measures the amount that space has expanded. By measuring how much the light from distant supernovae has redshifted, cosmologists can then infer how much smaller the universe was at the time of the explosion as compared with its size today. The larger the redshift, the smaller the universe was when the supernova occurred and hence the more the universe has expanded since then.

The observed brightness of a supernova provides a measure of its distance from us, which in turn reveals how much time has elapsed since it occurred. If a supernova with a given redshift looks dimmer than expected, then that supernova must be farther away than astronomers thought. Its light has taken longer to reach us, and hence the universe must have taken longer to grow to its current size [*see box on opposite page*]. Consequently, the expansion rate of the universe must have been slower in the past than previously expected. In fact, distant supernovae are dim enough that the expansion of the universe must have accelerated to have caught up with its current expansion rate.

This accelerating expansion is the big surprise that fired the current revolution in cosmology. Matter in the universe should tug at the fabric of spacetime, slowing down the expansion, but the supernova data suggest otherwise. If cosmologists accept the cosmological principle and assume that this acceleration happens everywhere, we are led to the conclusion that the universe must be permeated by an exotic form of energy, dark energy, that exerts a repulsive force.

Nothing meeting the description of dark energy appears in physicists' Standard Model of fundamental particles and forces. It is a substance that has not as yet been measured directly, has properties unlike anything we have ever seen and has an energy density some 10^{120} times less than we may have naively expected. Physicists have ideas for what it might be, but they remain speculative. In short, we are very much in the dark about dark energy. Researchers are working on a number of ambitious and expensive missions to find and characterize dark energy. To many, it is the greatest challenge facing modern cosmology.

A LIGHTER ALTERNATIVE

CONFRONTED WITH SOMETHING SO strange and seemingly so improbable, some researchers are revisiting the reasoning that

Copernicus's Legacy

The Copernican principle holds that Earth does not occupy a special place in the universe. The universe has a uniform density (homogeneity) and looks the same in every direction (isotropy).

Though powerful, the principle applies only on scales much larger than a galaxy. After all, if the cosmos were completely uniform, it would be a thin gruel of atoms rather than a constellation of galaxies. Also, the principle applies in space but not in time. We live in a special era—long enough after the big bang that complex life can form but not so long that stars have all died off.

Copernicus is commonly associated with a dethroning of humanity from any position of importance. But as historian Dennis Danielson of the University of British Columbia argues, although pre-Copernican Europeans placed Earth at the center of the universe, they did not consider the center a position of importance but quite the opposite—as Galileo put it, "the sump where the universe's filth and ephemera collect."

IN BRIEF

The universe appears to be expanding at an accelerating rate, implying the existence of a strange new form of energy—dark energy. The problem: no one is sure what dark energy is.

Cosmologists may not actually need to invoke exotic forms of energy. If we live in an emptier-than-average region of space, then the cosmic expansion rate varies with position.

A giant void strikes most cosmologists as highly unlikely, but so for that matter does dark energy. Observations over the coming years will differentiate between the two possibilities.

GETTY IMAGES (*Copernicus*)

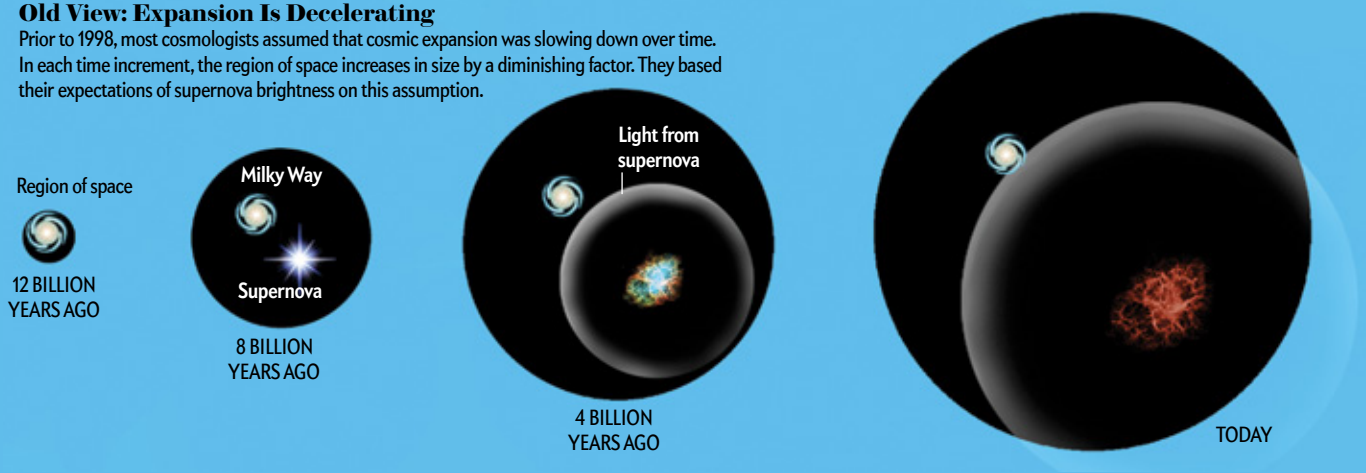
© 2013 Scientific American

Three Ways of Expanding a Universe

Astronomers have found that distant supernovae explosions are dimmer than expected. To see what this discovery means for cosmic expansion, consider a region of space that encompasses a supernova and our Milky Way galaxy. Over time this region gets bigger as the fabric of space stretches like a rubber sheet. The supernova goes off when the universe is about half its current size (which occurs at different times depending on whether the expansion is decelerating or accelerating). Light from the explosion spreads out, eventually reaching us on the outskirts of the Milky Way.

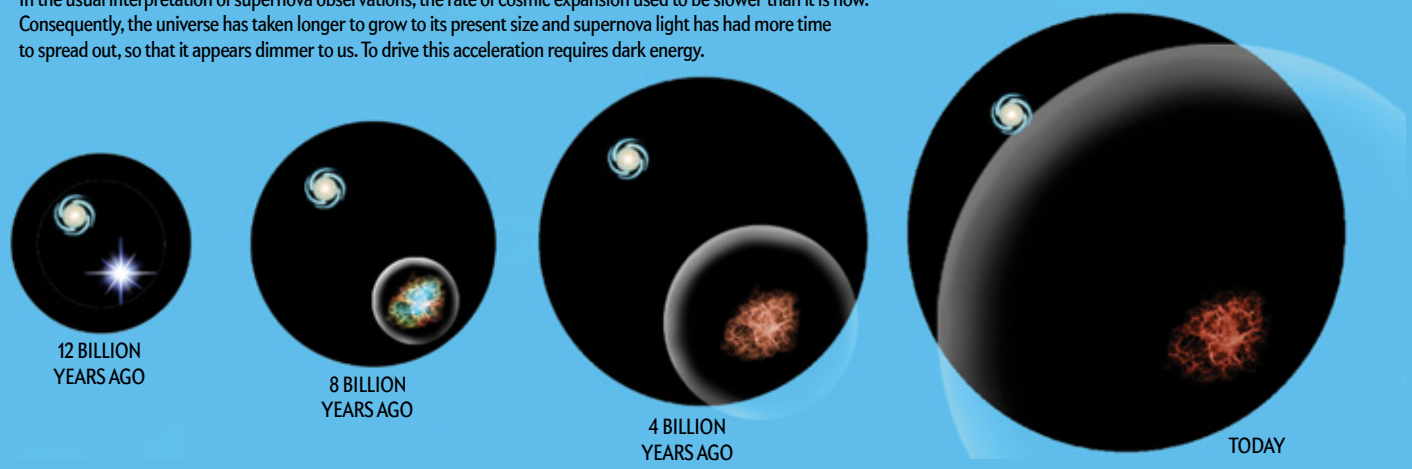
Old View: Expansion Is Decelerating

Prior to 1998, most cosmologists assumed that cosmic expansion was slowing down over time. In each time increment, the region of space increases in size by a diminishing factor. They based their expectations of supernova brightness on this assumption.



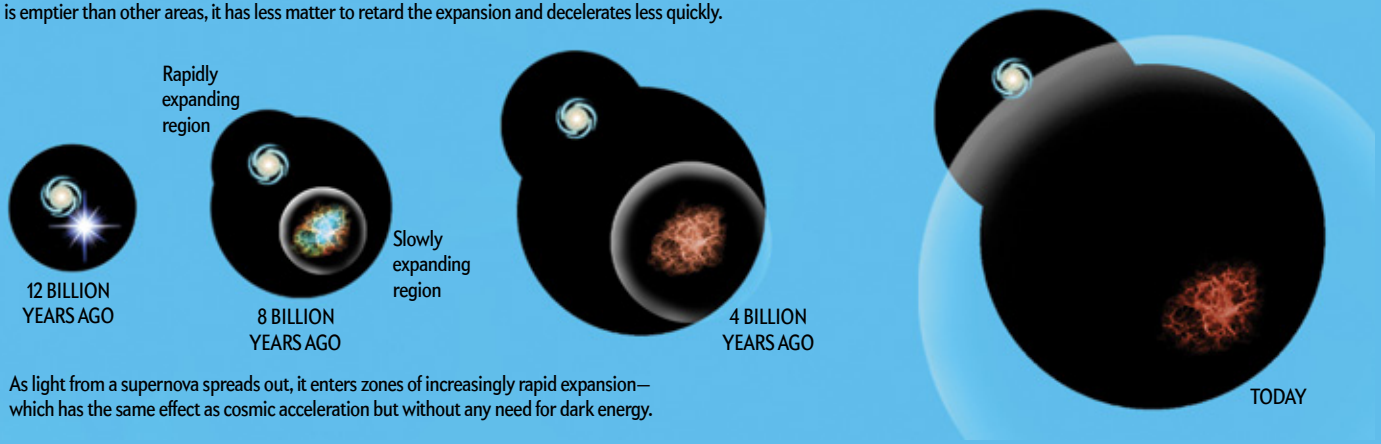
Scenario 1: Expansion Is Accelerating

In the usual interpretation of supernova observations, the rate of cosmic expansion used to be slower than it is now. Consequently, the universe has taken longer to grow to its present size and supernova light has had more time to spread out, so that it appears dimmer to us. To drive this acceleration requires dark energy.



Scenario 2: Universe Is Inhomogeneous

Alternatively, perhaps expansion is decelerating but at different rates in different places. If our neighborhood is emptier than other areas, it has less matter to retard the expansion and decelerates less quickly.



As light from a supernova spreads out, it enters zones of increasingly rapid expansion—which has the same effect as cosmic acceleration but without any need for dark energy.

© 2013 Scientific American

led them to it. One of the primary assumptions they are questioning is whether we live in a representative part of the universe. In the conventional picture, we talk about the expansion of the universe on the whole. It is very much like when we talk about a balloon blowing up: we discuss how big the entire balloon gets, not how much each individual patch of the balloon inflates. But we all have had experience with those annoying party balloons that inflate unevenly. One ring stretches quickly, and the end takes a while to catch up. In an alternative view of the universe, one that jettisons the cosmological principle, space, too, expands unevenly. A more complex picture of the cosmos emerges.

Consider the following scenario, first suggested by George Ellis and Charles Hellaby of the University of Cape Town in South Africa and Nazeem Mustapha, now at South Africa's Human Sciences Research Council, and subsequently followed up by Marie-Noëlle Célérier of the Paris-Meudon Observatory in France. Suppose that the expansion rate is decelerating everywhere, as matter tugs on spacetime and slows it down. Suppose, further, that we live in a gargantuan cosmic void—not a completely empty region, but one in which the average density of matter is only a half or maybe a third of the density elsewhere. The emptier a patch of space is, the less matter it contains to slow down the expansion of space; accordingly, the local expansion rate is faster within the void than it is elsewhere. The expansion rate is fastest at the center of the void and diminishes toward the edge, where the higher-density exterior begins to make itself felt. At any given time, different parts of space will expand at different rates, like the unevenly inflated party balloon.

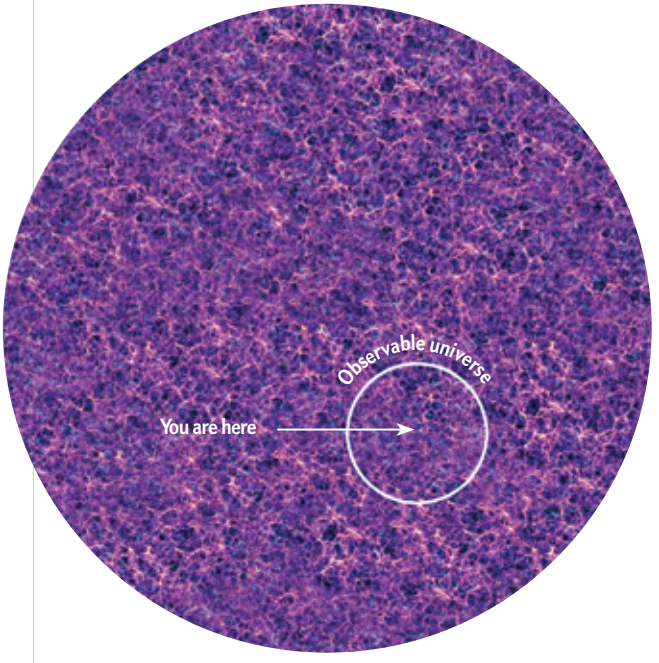
Now imagine supernovae exploding in different parts of this inhomogeneous universe, some close to the center of the void, others nearer the edge and some outside the void. If we are near the center of the void and a supernova is farther out, space expands faster in our vicinity than it does at the location of the supernova. As light from the supernova travels toward us, it passes through regions that are expanding at ever faster rates. Each region stretches the light by a certain amount as it passes through, and the cumulative effect produces the redshift we observe.

Light traveling a given distance is redshifted by less than it would be if the whole universe expanded at our local rate. Conversely, to achieve a certain redshift in such a universe, the light has to travel a greater distance than it would in a uniformly expanding universe, in which

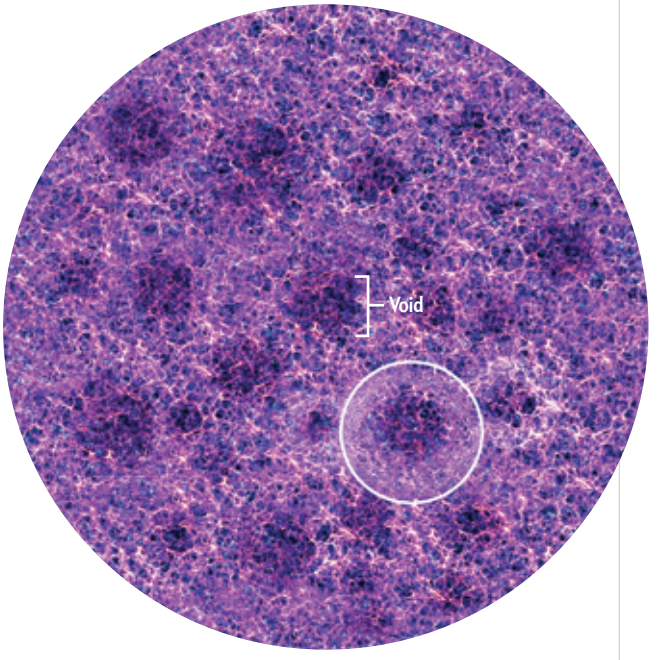
THE BIG PICTURE

A Special Place for Us

In his *Hitchhiker's Guide to the Galaxy* series of novels, Douglas Adams imagines a torture device that drives people insane by showing them the utter insignificance of their place in the universe. One would-be victim emerges unscathed when it turns out that the universe does, in fact, revolve around him. In a case of life imitating art, many cosmologists are investigating whether our planet indeed has a special place within the grand scheme of things.



Homogeneous Universe: Our Location Is Typical
In the standard view, galaxies are lined up in a spidery pattern, but overall space looks much the same everywhere, and Earth's position is nothing special.



Inhomogeneous Universe: Our Location Is Special
Alternatively, the density of matter could vary on large scales, and Earth may lie at or near the center of a relatively less dense region, or void.

© 2013 Scientific American

case the supernova has to be farther away and therefore appear dimmer.

Another way to put it is that a variation of expansion rate with position mimics a variation in time. In this way, cosmologists can explain the unexpected supernova observations without invoking dark energy. For such an alternative explanation to work, we would have to live in a void of truly cosmic proportions. The supernova observations extend out to billions of light-years, a significant fraction of the entire observable universe. A void would have to be of similar size. Enormous by (almost) anyone's standards.

A FAR-FETCHED POSSIBILITY

SO HOW OUTLANDISH is this cosmic void? At first glance, very. It would seem to fly in the face of the cosmic microwave background, which is uniform to one part in 100,000, not to mention the apparently uniform distribution of galaxies. On closer inspection, however, this evidence may not be so conclusive.

The uniformity of the relic radiation merely requires the universe to look nearly the same in every direction. If a void is roughly spherical and if we lie reasonably close to its center, these observations do not necessarily preclude it. In addition, the cosmic microwave background has some anomalous features that could potentially be explained by large-scale inhomogeneity [*see box on next page*].

As for the galaxy distribution, existing surveys do not extend far enough to rule out a void of the size that would mimic dark energy. They identify smaller voids, filaments of matter and other structures hundreds of millions of light-years in size, but the putative void is an order of magnitude larger. An analysis by David Hogg of New York University and his collaborators indicates that the largest structures in the universe are about 200 million light-years in size; on larger scales, matter appears smoothly distributed, in accordance with the cosmological principle. But Francesco Sylos Labini of the Enrico Fermi Center in Rome and his colleagues argue that the largest structures discovered so far are limited only by the size of the galaxy surveys that found them. Still larger structures might stretch beyond the scope of the surveys.

By analogy, suppose you had a map showing a region 10 miles wide, on which a road stretched from one side to the other. It would be a mistake to conclude that the longest possible road is 10 miles long. To determine the length of the longest road, you would need a map that clearly showed the end points of all roads, so that you would know their full extent. Similarly, astronomers need a galaxy survey that is larger than the biggest structures in the universe if they are to prove the cosmological principle. Whether surveys are big enough yet is the subject of a lively debate.

For theorists, too, a colossal void is difficult to stomach. All available evidence suggests that galaxies and larger structures such as filaments and voids grew from microscopic quantum seeds that cosmic expansion enlarged to astronomical proportions, and cosmological theory makes firm predictions for how many structures should exist with a certain size. The larger a structure is, the rarer it should be. The probability of a void big enough to mimic dark energy is less than one part in 10^{100}. Giant voids may well exist out there, but the chance of our finding one in our observable universe would seem to be tiny.

Still, there is a possible loophole. In the early 1990s one of the authors of what is now the standard model of the early universe, Andrei Linde, and his collaborators at Stanford University showed that although giant voids are rare, they expand faster early on and come to dominate the volume of the universe. The probability of observers finding themselves in such a structure may not be so tiny after all. This result shows that the cosmological principle (that we do not live in a special place) is not always the same thing as the principle of mediocrity (that we are typical observers). One can, it seems, be both typical and live in a special place.

TESTING THE VOID

WHAT OBSERVATIONS could tell whether the expansion of the universe is driven by dark energy or whether we are living in a special place, such as at the center of a giant void? To test for the presence of a void, cosmologists need a working model of how space, time and matter should behave in its vicinity. Just such a model was formulated in 1933 by Abbé Georges Lemaître, independently rediscovered a year later by Richard Tolman and further developed after World War II by Hermann Bondi (who died in 2005). The universe they envisaged had expansion rates that depended not only on time but also on distance from a specific point, just as we now hypothesize.

With the Lemaître-Tolman-Bondi model in hand, cosmologists can make predictions for a range of observable quantities. To begin, consider the supernovae that first led to the inference of dark energy. The more supernovae that astronomers observe, the more accurately they can reconstruct the expansion history of the universe. Strictly speaking, these observations cannot ever rule out the void model, because cosmologists could re-create any set of supernova data by choosing a suitably shaped void. Yet for a void to be completely indistinguishable from dark energy, it would have to have some very strange properties indeed.

The reason is that the putative accelerating expansion occurs right up to the present moment. For a void to mimic it exactly, the expansion rate must decrease

No A-Voiding It

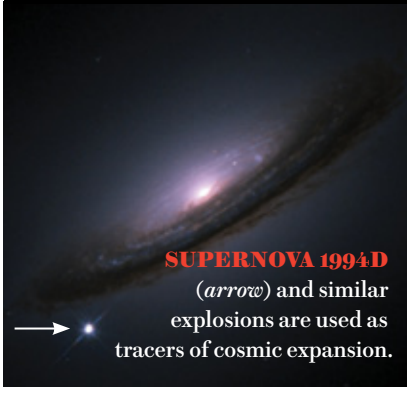
Although a cosmic void mimics dark energy, the match is not exact. Observations can look for telltale differences.

Additional supernova observations will pin down the expansion rate and check whether it varies with position, as a void model predicts.

Galaxy clusters reflect light and, in effect, let us view our cosmic neighborhood in the mirror. If we live in a void, we should be able to see it.

Galaxies and galaxy clusters evolve at a pace that depends on the expansion rate at their location and therefore on the presence of a void.

Neutrinos left over from the primordial universe could reveal a void.



SUPERNOVA 1994D (*arrow*) and similar explosions are used as tracers of cosmic expansion.

COURTESY OF NASA/ESA, HUBBLE KEY PROJECT TEAM AND HIGH-Z SUPERNOVA SEARCH TEAM

© 2013 Scientific American

sharply away from us and in every direction. Therefore, the density of matter and energy must increase sharply away from us in every direction. The density profile must look like an upside-down witch's hat, the tip of which corresponds to where we live. Such a profile would go against all our experience of what structures in the universe look like: they are known to be usually smooth, not pointy. Even worse, Ali Vanderveld and Éanna Flanagan, both then at Cornell University, showed that the tip of the hat, where we live, would have to be a singularity, like the ultradense region at the center of a black hole.

If, however, the void has a more realistic, smooth density profile, then a distinct observational signature presents itself. Smooth voids still produce observations that could be mistaken for acceleration, but their lack of pointyness means that they do not reproduce *exactly* the same results as dark energy. In particular, the apparent rate of acceleration varies with redshift in a telltale way. In a paper with Kate Land, then at the University of Oxford, we showed that several hundred new supernovae, on top of the few hundred we currently have, should be enough to settle the issue. Supernova-observing missions stand a very good chance of achieving this goal soon.

Supernovae are not the only observables available. In 1995 Jeremy Goodman of Princeton University suggested another possible test using the microwave background radiation. At the time, the best evidence for dark energy had not yet emerged, and Goodman was not seeking an explanation for any unexplained phenomena but proof of the Copernican principle itself. His idea was to use distant clusters of galaxies as mirrors to look at the universe from different positions, like a celestial dressing room. Galaxy clusters reflect a small fraction of the microwave radiation that hits them. By carefully measuring the spectrum of this radiation, cosmologists could infer some aspects of what the universe would look like if viewed from one of them. If a shift of viewpoint changed how the universe looked, it would be powerful evidence for a void or a similar structure.

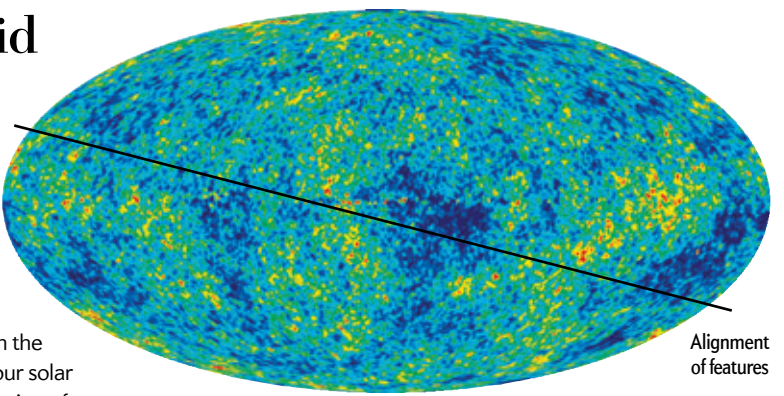
Two teams of cosmologists later put this idea to the test. Robert Caldwell of Dartmouth College and Albert Stebbins of the Fermi National Accelerator Laboratory in Batavia, Ill., studied precise measurements of distortions in the microwave background, and Juan García-Bellido of the Autonomous University of Madrid and Troels Haugbølle, then at Århus University in Denmark, looked at individual clusters directly. Neither group detected a void; the best the researchers could do was to narrow down the properties that such a void could have. The Planck satellite, launched in May 2009, should be able to place stronger limits on the void properties and maybe rule out a void altogether.

A third approach, advocated by Bruce Bassett, Chris Clarkson and Teresa Lu, all at the University of Cape Town, is to make independent measurements of the expansion rate at different locations. Astronomers usually measure expansion rates in terms of redshift, which is the cumulative effect of the expansion of all regions of space between a celestial body and us. By lumping all these regions together, redshift cannot distinguish a variation of expansion rate in space from a variation in time. It would be better to measure the expansion rate at specific

ASYMMETRY

Surrender to the Void

Most suggestions that we live in a cosmic void place us at its center, but what if we lived away from the middle? The universe would then look slightly lopsided. Håvard Alnes and Morad Amarzguioui, both then at the University of Oslo, have shown that the cosmic microwave background radiation would look slightly hotter in one direction than in the other. Such an asymmetry, called a dipole, has indeed been observed in the microwave background. It is usually attributed to our solar system's motion through space but could also be a sign of a lumpy universe.



Furthermore, small fluctuations in the microwave background appear to align in a specific direction—dubbed the "axis of evil" by João Magueijo of Imperial College London and Kate Land, then at Imperial College London as well. This alignment picks out a preferred direction in the sky, which, though hard to imagine in a Copernican universe, might be explained in terms of our displacement from the center of a void. A preferred direction would also have other effects, such as large-scale coherent motions of galaxies and galaxy clusters. Several researchers have claimed to have detected such a "dark flow," but it remains controversial.

Although it is tempting to attribute these anomalies to a giant void, this explanation does not really hold together. For a start, these effects each pick out different directions. Furthermore, the strength of the cosmic dipole would suggest that we are only about 50 million light-years from the center, which is only a very small fraction of the total size of the putative void.

—T.C. and P.G.F.

AXIS OF EVIL, an alignment of features in the cosmic microwave background radiation, could be a sign that we live in an inhomogeneous universe.

COURTESY OF NASA/WMAP SCIENCE TEAM

© 2013 Scientific American

spatial locations, separating out the effects of expansion at other locations. That is a difficult proposition, though, and has yet to be done. One possibility is to observe how structures form at different places. The formation and evolution of galaxies and galaxy clusters depend, in large part, on the local rate of expansion. By studying these objects at different locations and accounting for other effects that play a role in their evolution, astronomers may be able to map out subtle differences in expansion rate.

A NOT SO SPECIAL PLACE

THE POSSIBILITY that we live in the middle of a giant cosmic void is an extreme rejection of the cosmological principle, but there are gentler possibilities. The universe could obey the cosmological principle on large scales, but the smaller voids and filaments that galaxy surveys have discovered might collectively mimic the effects of dark energy. Tirthabir Biswas and Alessio Notari, both then at McGill University, as well as Valerio Marra and his collaborators, then at the University of Padua in Italy and the University of Chicago, have studied this idea. In their models, the universe looks like Swiss cheese—uniform on the whole but riddled with holes. Consequently, the expansion rate varies slightly from place to place. Rays of light emitted by distant supernovae travel through a multitude of these small voids before reaching us, and the variations in the expansion rate tweak their brightness and redshift. So far, however, the idea does not look very promising. One of us (Clifton), together with Joseph Zuntz of Oxford, showed that reproducing the effects of dark energy would take lots of voids of very low density, distributed in a special way.

Another possibility is that dark energy is an artifact of the mathematical approximations that cosmologists routinely use. To calculate the cosmic expansion rate, we typically count up how much matter a region of space contains, divide by the volume of the region and arrive at the average energy density. We then insert this average density into Einstein's equations for gravity and determine the averaged expansion rate of the universe. Although the density varies from place to place, we treat this scatter as small fluctuations about the overall average.

The problem is that solving Einstein's equations for an averaged matter distribution is *not* the same as solving for the real matter distribution and then averaging the resulting geometry. In other words, we average and then solve, when really we should solve and then average.

Solving the full set of equations for anything even vaguely approximating the real universe is unthinkably difficult, and so most of us resort to the simpler route. Thomas Buchert of the University of Lyon in France has taken up the task of determining how good an approximation it really is. He has introduced an extra set of terms into the cosmological equations to account for the error introduced by averaging before solving.

If these terms prove to be small, then the approximation is good; if they are large, it is not. The results so far are inconclusive. Some researchers have suggested that the extra terms may be enough to account for dark energy entirely, whereas others claim that they are negligible.

Observational tests may help distinguish between dark energy and the void models. The Planck spacecraft and a variety of ground-based and balloon-borne instruments are mapping out the microwave background in ever greater detail. The Square Kilometer Array, a gigantic radio telescope planned for 2020, will supply us with a survey of all the galaxies within our observable horizon.

Since the first publication of this article in April 2009, our ability to test the Copernican principle has progressed significantly. Together with our Oxford colleague Philip Bull, we have been able to take Goodman's idea about using clusters of galaxies as mirrors and combine it with observations of supernovae and the cosmic microwave background to form a test of the void hypothesis. Although a giant void can explain each of these observations individually, it cannot explain them all at once in a consistent way.

So around 500 years after Copernicus suggested it, we finally have some proof that we do not live at a central point in the universe. This is not the end of the story, however, as researchers around the world are continuing to work at constructing even more sophisticated models of the universe and tests for determining its geometry. SA

Walking the Planck

The latest spacecraft to measure the cosmic microwave background radiation, the European Space Agency's Planck, launched in May 2009.

Planck recently delivered its first all-sky inventory of fluctuations in the temperature of the microwave background, thereby extending an observational effort begun in the 1960s. These fluctuations reveal what the universe looked like at the tender age of 380,000 years and how it has grown since. The latest data include subtle anomalies that have yet to be explained.

Planck is also measuring fluctuations in the polarization (or directionality) of the radiation, which could reveal whether gravitational waves coursed through the ancient universe as a result of high-energy processes a fraction of a second after the big bang—or even before it.

COURTESY OF ESA/AOES MEDIALAB

Timothy Clifton and **Pedro G. Ferreira** are cosmologists at Queen Mary, University of London, and the University of Oxford, respectively. Both study the physics of the early universe and potential modifications to Einstein's general theory of relativity. Clifton, a keen oenophile, says his true interest in life is Burgundy wine. Ferreira is author of a popular-level astronomy book, *The State of the Universe,* runs a program for artists in residence at Oxford, and participates in various projects to support science education in Africa.

MORE TO EXPLORE

Geocentrism Reexamined. Jeremy Goodman in *Physical Review D,* Vol. 52, No. 4, pages 1821–1827; March 15, 1995. http://arxiv.org/abs/astro-ph/9506068

The State of the Universe: A Primer in Modern Cosmology. Pedro G. Ferreira. Phoenix, 2007.

Cosmology: Patchy Solutions. G.F.R. Ellis in *Nature,* Vol. 452, pages 158–161; March 12, 2008.

Living in a Void: Testing the Copernican Principle with Distant Supernovae. Timothy Clifton, Pedro G. Ferreira and Kate Land in *Physical Review Letters,* Vol. 101, Article No. 131302; September 26, 2008. http://arxiv.org/abs/0807.1443

© 2013 Scientific American

ASTRONOMY

PORTRAIT OF A BLACK HOLE

By adapting a global network of telescopes, astronomers will soon get their first look ever at the dark silhouette of a black hole

By Avery E. Broderick and Abraham Loeb

IN BRIEF

Black holes are among the most mysterious objects in the universe. So far astronomers have observed them only indirectly, from their gravitational effects on stars and from the radiation emitted by hot gas spiraling toward them.

Astronomers are adapting a network of radio telescopes to produce images of the supermassive black holes that lie at the center of the Milky Way and M87 galaxies.

Better studies of black holes not only would help explain the unusual phenomena that are generated by the holes but also could test Einstein's general theory of relativity and begin to provide vital insights into the nature of gravity in extreme situations.

© 2013 Scientific American

YOU HAVE PROBABLY seen the television commercial in which a cell phone technician travels to remote places and asks on his phone, "Can you hear me now?" Imagine this technician traveling to the center of our Milky Way galaxy, wherein lurks a massive black hole, Sagittarius A* (Sgr A*), weighing as much as 4.5 million suns. As the technician approached within 10 million kilometers of the black hole, we would hear his cadence slow down and his voice deepen and fade, eventually turning to a monotone whisper with diminishing reception. If we were to look, we would see his image turn increasingly red and dim as he became frozen in time near the black hole's boundary, known as the event horizon.

The technician himself, however, would experience no slowing of time and would see nothing strange at the location of the event horizon. He would know he had crossed the horizon only when he heard us say, "No, we cannot hear you very well!" He would have no way of sharing his last impressions with us—nothing, not even light, can escape from gravity's extreme pull inside the event horizon. A minute after he crossed the horizon, the gravitational forces deep inside the hole would tear him apart.

In real life we cannot send a technician on such a journey. But astronomers have developed techniques that will soon allow them, for the first time, to produce images of a black hole's dark silhouette against a backdrop of hot glowing gas.

Wait, you say. Haven't astronomers reported lots of observations of black holes, including all sorts of pictures? That is true, but the pictures have been of gas or other material near a black hole, with the hole itself an invisible speck, or of huge outpourings of energy presumed to come from a black hole. In fact, we do not even know for certain whether black holes really exist [see "Black Stars, Not Holes," by Carlos Barceló, Stefano Liberati, Sebastiano Sonego and Matt Visser, on page 82].

MENACINGLY DARK DISK of the Milky Way galaxy's central black hole and a hotspot in the gas orbiting the hole are both consistent with computer simulations (*opposite page* and *above left*) based on telescope observations made over the past five years. Interstellar gas does, however, blur the finer details (*above right*).

COURTESY OF AVERY E. BRODERICK

© 2013 Scientific American

Astronomers have detected objects in the sky that are sufficiently massive and compact that, if Einstein's general theory of relativity is correct, they must be black holes, and it is customary to talk of them as if they were (as we do in this article). But until now we could not tell if these objects had the defining characteristic of a black hole—a horizon through which material can flow only one way. This question is not merely a matter of esoteric curiosity, because such horizons are at the heart of one of the deepest puzzles in theoretical physics. And images showing the dark silhouettes of black holes' event horizons would help us understand the extraordinary astrophysical processes taking place in their neighborhood.

DRIVING QUESTIONS

EVENT HORIZONS are a source of fascination because they represent a fundamental inconsistency between two great triumphs of 20th-century physics: quantum mechanics and general relativity. Time reversibility is an essential feature of the quantum-mechanical description of physical systems; every quantum process has an inverse process, which may be used, in principle, to recover any information that the original process may have scrambled. In contrast, general relativity—which explains gravity as arising from the curvature of space and predicts the existence of black holes—admits no inverse process to bring back something that has fallen into a black hole.

The need to resolve this inconsistency between quantum mechanics and gravitation has been one of string theorists' primary motivations in their quest for a quantum theory of gravity—a theory that would predict the properties of gravitation as arising from interactions following the laws of quantum mechanics.

On a more basic level, physicists would like to know if Einstein's general relativity really is *the* theory of gravity, even where it predicts shocking deviations from classical, Newtonian theory—such as the existence of event horizons. Black holes have the twin virtues of corresponding to extraordinarily simple solutions to Einstein's equations of gravity (a black hole is completely characterized by just three numbers—its mass, charge and spin), as well as being places where gravity differs the most from Newtonian theory. Thus, black holes are prime locations for seeking evidence of deviations from Einstein's equations under extreme conditions, which could provide clues toward a quantum theory of gravity. Conversely, the equations' success near black holes will dramatically extend the regime in which we know general relativity works.

Pressing astrophysical questions about what happens in the vicinity of black holes also demand answers. Black holes are fed by infalling material such as gas and dust. The matter gains vast quantities of energy as it falls closer to the hole's horizon, producing heat 20 times more efficiently than nuclear fusion, the next most potent energy generator known. Radiation from the hot, spiraling gas makes the environment near black holes the brightest objects in the universe.

Astrophysicists can model the accreting material to some extent, but it is unclear how gas in the accretion flow migrates from an orbit at a large radius to one near the horizon and how, precisely, it finally falls into the black hole. Magnetic fields, created by charged particles moving in the accretion flow, must play a very important role in how the flow behaves. Yet we know little about how these fields are structured and how that structure affects black holes' observed properties. Although computer simulations of the entire accreting region are becoming feasible, we theorists remain decades away from true ab initio calculations. Input from observations will be vital for inspiring new ideas and deciding among competing models.

Sgr A* is the only supermassive hole close enough that telescopes can resolve the individual stars nearest to it.

More embarrassing to astrophysicists is our lack of understanding of black hole jets—phenomena in which the forces near a supermassive black hole somehow conspire to spew out material at ultrarelativistic speeds (up to 99.98 percent of light speed). These amazing outflows traverse distances larger than galaxies, yet they originate near the black hole as intense beams collimated tightly enough that they could thread the solar system—the eye of a galactic needle. We do not know what accelerates these jets to such high speeds or even what the jets are made of—are they electrons and protons or electrons and positrons, or are they primarily electromagnetic fields? To answer these and other questions, astronomers desperately need direct observations of the gas in a black hole's vicinity.

STALKING THE BEHEMOTH FROM AFAR

UNFORTUNATELY, SUCH OBSERVATIONS are difficult for several reasons. First, black holes are extremely small by any astronomical measure. They appear to come in two main varieties: stellar-mass black holes, the remnants of dead massive stars, with typical masses of five to 15 suns, and supermassive black holes, located at the centers of galaxies and weighing millions to 10 billion suns. A 15-solar-mass black hole's event horizon would be a mere 90 kilometers in diameter—far too tiny to be resolved at interstellar distances. Even a one-billion-sun monster would fit comfortably inside Neptune's orbit.

Second, a black hole's small size and intense gravity make for extremely fast motion—matter very near a stellar-mass black hole can complete an orbit in less than a millisecond. It takes highly sensitive instruments to observe such rapid phenomena. Finally, only the small subset of black holes that have large reservoirs of nearby gas to accrete are visible at all; the vast majority of black holes in the Milky Way are, as yet, undiscovered.

Rising to these challenges, astronomers have developed a variety of techniques that, short of providing direct images, have provided information about the properties and behavior of matter orbiting close to suspected black holes. For instance, astronomers can weigh a supermassive black hole by observing stars near it, much like using planets' orbits to weigh the sun. In distant galaxies, individual stars near a supermassive hole cannot be resolved, but the spectrum of their light indicates their distribution of velocities, which yields a mass for the hole. The supermassive black hole Sgr A* at the center of the Milky Way is close enough for

© 2013 Scientific American

telescopes to resolve individual stars near it, producing the best mass estimate of any black hole to date [*see box on next page*]. Unfortunately, these stars are far outside the region that interests us most, where general relativistic effects become significant.

Astronomers also search for signatures of general relativity in the way that radiation emitted near a black hole varies over time. For example, the x-ray emissions of some stellar-mass black holes fluctuate in luminosity in a nearly periodic manner with periods similar to that of orbits expected to be near the inner edge of the accretion disk.

Thus far the most fruitful avenue for probing supermassive black holes has exploited the fluorescence of iron atoms on the surface of the accretion disk. The fast motion of the accretion disk carrying the iron atoms and the strong gravity of the hole combine to shift the characteristic wavelength of the fluorescence, spreading it over a band of wavelengths. Near a rapidly spinning black hole, the accretion disk itself orbits the hole faster (thanks to a general relativistic effect that drags space around with the hole's rotation), and the emission will have a telltale asymmetry. The Japanese satellites ASCA and Suzaku have observed just such emissions, which astronomers interpret as direct evidence of rapidly spinning black holes, with orbital velocities as high as one third of light speed in the accretion disks.

Information about how much spin stellar-mass black holes have has come from binary systems in which a black hole and an ordinary star orbit each other close enough for the hole to slowly feed on its companion. Analysis of the x-ray spectra and orbital parameters for a handful of such systems indicates that the holes have 65 to 100 percent of the maximum spin permitted by general relativity for a hole of a given mass; very high spin seems to be the norm.

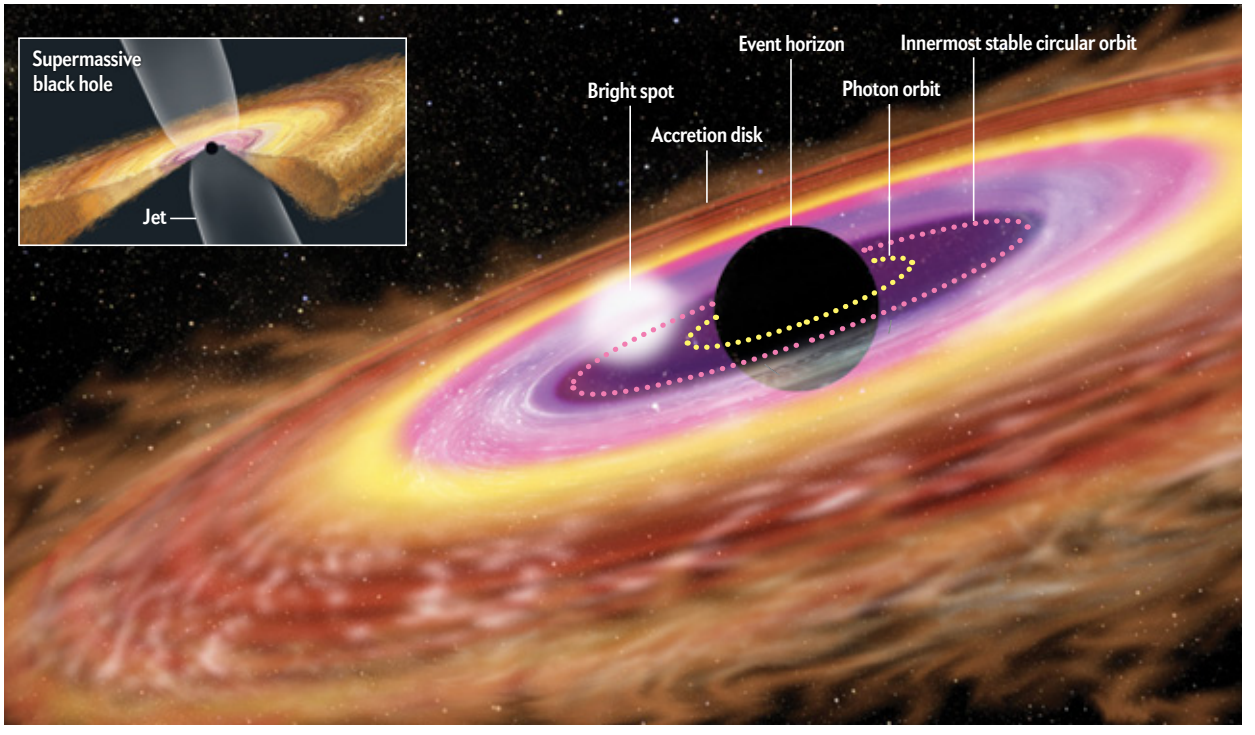
Light (ranging from radio waves to x-rays) and energetic jets are not the only things emitted by black holes. When two black holes collide, they shake the fabric of spacetime around them, producing gravitational waves that propagate out like ripples on a pond. These ripples of spacetime should be detectable at vast distances, albeit requiring incredibly sensitive instruments. Although observatories already operating have yet to detect any gravitational waves, the method offers a revolutionary new way to study black holes.

BASICS

Lair of a Monster

A black hole's defining feature is its event horizon, the spherical boundary of the region within which nothing can overcome the black hole's gravity and emerge. Gas accretes in a hot, luminous disk orbiting the black hole, with transitory bright spots similar to solar flares. A disk may be thin, as depicted, but can also span a large angle above and below the plane of rotation, as well as extending much farther radially. Many supermassive black holes emit bright jets at almost the speed of light.

The accretion disk's inner edge is believed to be near a circle called the innermost stable circular orbit. Any matter that strays closer to the hole will find itself in an unstable orbit and quickly plunge into the hole. At the photon orbit, light could in principle circle the black hole permanently, but in practice the tiniest disturbance would send the light spiraling in or out.



JEN CHRISTIANSEN (*inset*)

Illustration by Don Dixon

© 2013 Scientific American

A WINDOW WITH A VIEW

DESPITE PROVIDING a wealth of information, none of the techniques we have described thus far offer an image of a black hole's event horizon. Now, however, thanks to very recent advances in technology, direct imaging of a black hole's horizon is imminent. The black hole to be imaged is the behemoth in our backyard, Sgr A*. At a distance of only 24,000 light-years, Sgr A* occupies the largest disk on the sky of any known black hole. A 10-solar-mass black hole would have to be 1/100th as far away as the nearest star to appear as big. And although supermassive black holes much larger than Sgr A* exist, they are millions of light-years away.

The dark silhouette of a distant black hole is more than doubled in apparent size thanks to the bending of light rays by the hole's gravity. Even so, Sgr A*'s horizon will appear to span a mere 55 microarcseconds—as small as a poppy seed in Los Angeles viewed from New York City.

The resolution of all modern telescopes, as impressive as they are, is fundamentally limited by diffraction, a wave-optics effect that occurs as light passes through the finite aperture presented by the telescope's size. Generally, the smallest angular scale resolvable by a telescope can be decreased by making the telescope larger or by capturing shorter wavelength light. At infrared wavelengths (which, conveniently, pass through dust clouds that hide Sgr A* at visible wavelengths), an angular scale of 55 microarcseconds would require a telescope that was seven kilometers across.

The shorter wavelengths of visible or ultraviolet light would help reduce this gargantuan requirement somewhat but not by enough to be any less ridiculous. Considering longer wavelengths might seem pointless—millimeter radio waves, for instance, would require a telescope 5,000 kilometers across. But it just so happens that Earth-size radio telescopes are already in operation.

A technique called very long baseline interferometry (VLBI) combines the signals detected by an array of radio telescopes sprinkled around the globe to achieve the angular resolutions that an Earth-size radio dish would achieve. Two such arrays of telescopes have been operating for more than a decade—the Very Long Baseline Array (VLBA), with dishes in the U.S. as far afield as Hawaii and New Hampshire, and the European VLBI Network (EVN), with dishes in China, South Africa and Puerto Rico, as well as Europe. You may remember seeing a much smaller system, the Very Large Array in New Mexico, in movies such as *Contact* and *2010*.

Unfortunately, the VLBA and EVN are suitable only for radio wavelengths above 3.5 millimeters, corresponding to resolutions of at best 100 microarcseconds, too large to resolve Sgr A*'s horizon. Moreover, at these wavelengths, interstellar gas blurs the image of Sgr A*, just as dense fog blurs the streetlights overhead. The solution is to implement an interferometer at shorter wavelengths of millimeters and below.

These shorter wavelengths, however, suffer yet another problem: absorption by atmospheric water vapor. For this reason, millimeter and submillimeter telescopes are placed at the highest, driest sites available, such as atop Mauna Kea in Hawaii, in the Atacama Desert in Chile, and in Antarctica. When all is said and done, two useful windows generally remain open, at 1.3 millimeters and at 0.87 millimeter. An Earth-size array at these wavelengths would provide resolutions of about 26 and 17 microarcseconds, respectively, good enough to resolve the horizon of Sgr A*.

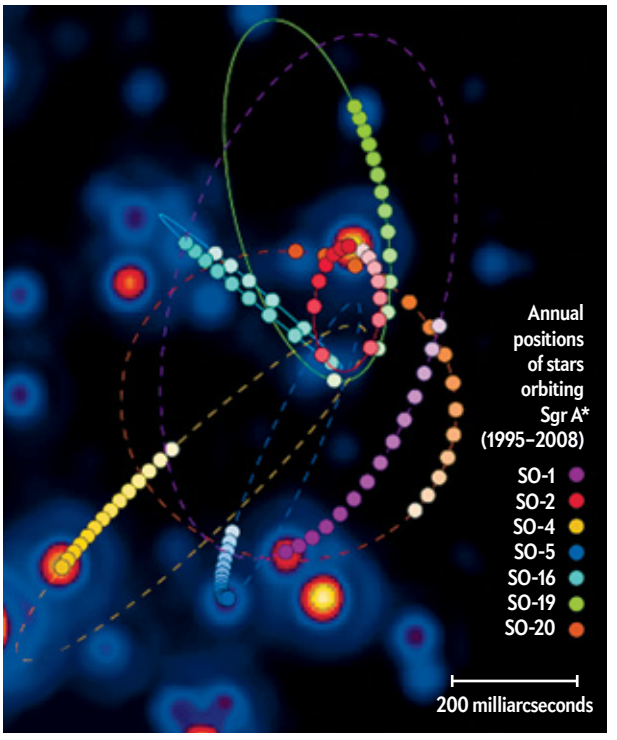
A number of millimeter and submillimeter telescopes that could be incorporated in such an array already exist—in Hawaii, scattered throughout the southwestern U.S., and in Chile, Mexico and Europe. Because astronomers built these telescopes for other purposes, adapting them for VLBI involves many technological challenges, including development of extraordinarily low-noise electronics and ultrahigh bandwidth digital recorders.

Nevertheless, a collaboration headed by Sheperd S. Doeleman of the Massachusetts Institute of Technology solved these problems in 2008. The group studied Sgr A* at 1.3-millimeter wavelengths with an array of just three telescopes, in Arizona, in California and on Mauna Kea. Such a small number of telescopes is insufficient to generate an image, but the researchers successfully

SAGITTARIUS A*

Closing in on the Giant

Until recently, observations of the motions of stars near the Milky Way's center have been the closest that astronomers have come to observing the event horizon of the black hole Sgr A*. The stars' orbits (*dashed lines*) reveal that they are in the thrall of a very compact object with a mass of 4.5 million suns. Colored dots mark the stars' positions each year from 1995 to 2008. The background is a 2008 infrared image of the stars (and others). The star S0-16 comes closest to Sgr A*—within seven light-hours—but even that distance remains 600 times larger than the event horizon's radius.

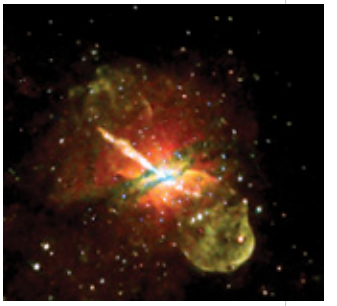


COURTESY OF ANDREA GHEZ *U.C.L.A. Galactic Center Group*; DATA SETS OBTAINED WITH THE W. M. KECK TELESCOPES

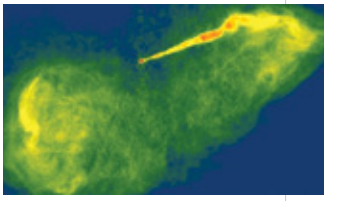
© 2013 Scientific American

Distant Signs of Black Holes

Astrophysicists believe that black holes—too small and far away to be seen—are responsible for phenomena ranging from x-ray emissions to huge jets of material ejected from galaxy centers.



The Centaurus A galaxy, imaged in x-rays by the Chandra satellite, shows a 13,000-light-year jet emitted by its suspected supermassive black hole (*top*). Star-like spots are stellar-mass holes consuming matter from companion stars. Red, green and blue indicate three bands of x-ray wavelengths.



A supermassive black hole presumed to be at the center of the M87 galaxy is surrounded by lobes of gas about 15,000 light-years across and shoots an ultrarelativistic jet toward us (*bottom*). An unseen counterjet may be sending material in the opposite direction. The Very Large Array in New Mexico took this image at two-centimeter radio wavelengths. The colors represent intensity of the signal.

SCIENCE SOURCE (*Centaurus A*); COURTESY OF NATIONAL RADIO ASTRONOMY OBSERVATORY/NATIONAL SCIENCE FOUNDATION (*M87*)

resolved Sgr A* in that their data indicated that it had bright regions only 37 microarcseconds in size, two thirds of the horizon's size. Additional telescopes will make it possible to produce images of the event horizon's dark silhouette. Collectively, they form the Event Horizon Telescope, with the three sites mentioned above as its nucleus.

Already the Event Horizon Telescope observations make it exceedingly difficult for Sgr A* not to have a horizon. Accretion onto a black hole and onto horizonless objects differ in a fundamental way. In both cases, the accreting material accrues vast amounts of energy during its infall. In the absence of a horizon, this energy is converted to heat where the accreting material finally settles and is subsequently emitted as radiation, producing a characteristic thermal spectrum visible to outside observers. In contrast, for black holes, infalling material can carry any amount of energy across the horizon, which will hide it forever.

For Sgr A*, we can use its total luminosity to estimate the infall rate of accreting material. The millimeter-VLBI observations place a tight limit on the maximum possible size of the accretion flow's inner edge and thus on how much energy has been liberated in the flow's fall to that point. If Sgr A* does not have a horizon (and so is not a black hole), the surplus energy must be radiated when the accreting material comes to rest, emitting primarily in the infrared. Despite careful observations, astronomers have not found any thermal infrared emission from Sgr A*. The only way to reconcile this discrepancy without a horizon is for the material to radiate away all the excess energy as it plummets inward, but that would require absurdly high radiative efficiencies.

PORTRAIT OF A MONSTER

WE, AMONG OTHER THEORISTS, have been very busy analyzing the first Event Horizon Telescope observations of Sgr A*, identifying clues to its fundamental nature. Generically, a black hole casts a silhouette on the wallpaper of emissions by nearby accreting gas. This "shadow" arises because the black hole swallows light rays coming toward the observer from just behind it. Meanwhile the bright region around the "shadow" is supplemented by other light waves from behind the black hole that just miss the horizon. Strong gravitational lensing bends light rays so that even material directly behind the black hole will contribute to the light around the dark region. The resulting silhouette is what is meant by a "portrait of a black hole," a fitting picture in which the black hole truly is black.

This shadow will not be a circular disk, primarily because of the extreme orbital velocities of the gas, which approach the speed of light. The emission from such fast-moving matter will be Doppler-shifted and concentrated in a narrow cone in the direction of motion, which substantially brightens the emission from the approaching side of the orbiting gas and dims the receding side, producing a bright crescent instead of a full, bright ring around a disklike silhouette. This asymmetry disappears only if we happen to be looking along the disk's axis of rotation.

The spin of the black hole itself, which may have a different axis of rotation than the accretion disk, has a similar effect. Such images will therefore allow astronomers to determine the direction of the black hole spin and the inclination of the accretion disk relative to it. Equally important for astrophysics, the data will provide invaluable observational input into accretion theory, settling once and for all the gas's density and the geometry of the accretion flow's inner edge.

Other supermassive black holes should also be within range of VLBI and can be compared with Sgr A*. A few years ago we showed that the second-best target is the black hole believed to lie at the center of the giant elliptical galaxy M87. This black hole is 55 million light-years distant, and its mass is estimated to be about 6.6 billion suns, giving it an expected silhouette comparable to the size of Sgr A*'s.

In many respects, M87 is a more interesting and promising target than Sgr A*. It has a vigorous jet that extends 5,000 light-years; resolving the jet-launching region will provide critical information for theorists' efforts to understand these ultrarelativistic outflows. Unlike Sgr A*, M87 lies in the Northern Hemisphere sky, making it more amenable to VLBI using existing observatories, relatively few of which are in the south. Furthermore, with the M87 black hole being 2,000 times the size of Sgr A*, dynamical changes will occur on timescales of days instead of minutes, and thus obtaining sequences of images of unfolding events should be much easier. (The orbital period near the inner edge of M87's accretion disk will be 0.5 to five weeks, depending on the hole's spin.) Finally, high-resolution images will most likely suffer less blurring of the kind inflicted by interstellar gas between us and Sgr A*. Already the latest VLBI data on M87, published at the end of 2012, reveal that the base of the jet is so compact (only 40 microarcseconds in width) that it probably originates from a spinning black hole. This is the

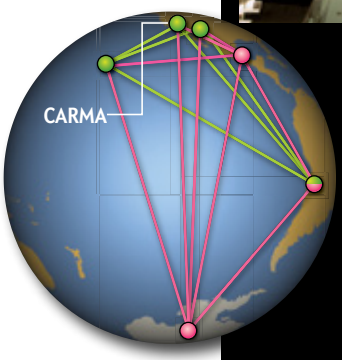
© 2013 Scientific American

IMAGING

Shooting the Beast

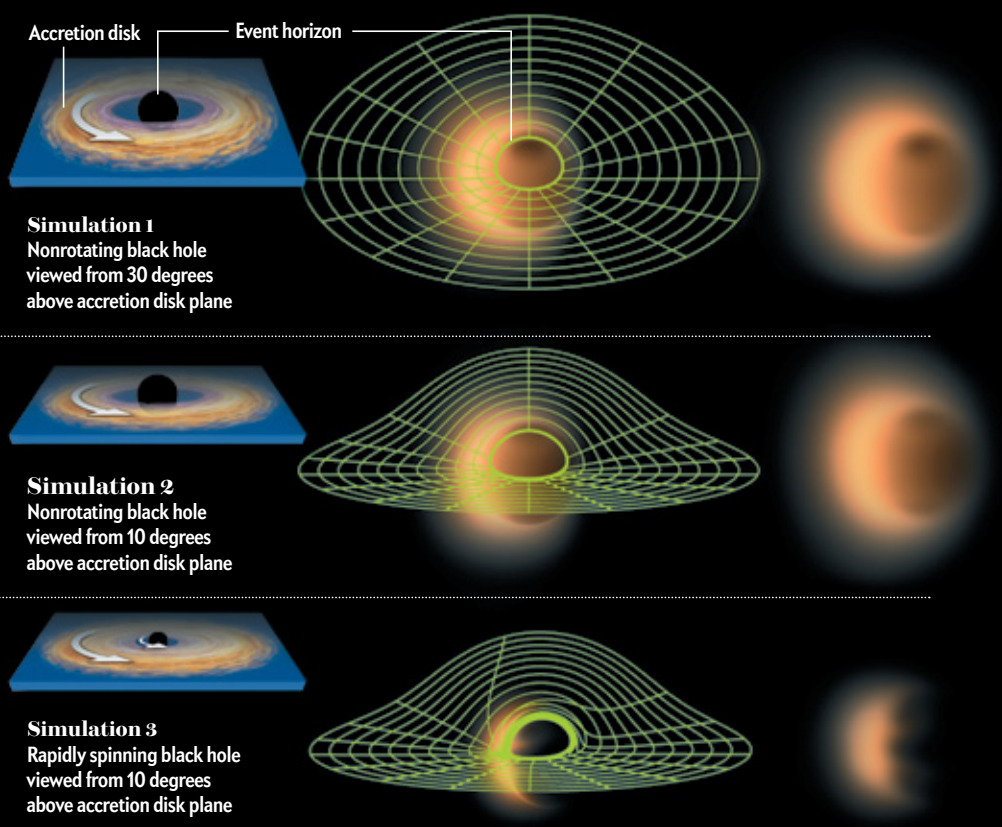
Astronomers have been developing several radio telescope arrays to form a globe-spanning network of observatories (*right*) that can observe Sgr A* and its immediate surroundings at wavelengths near 0.87 and 1.3 millimeters—two "windows" that are not excessively absorbed by Earth's atmosphere or scattered by interstellar gas. The size of the network will permit observations with sufficient resolution to produce images of Sgr A*'s event horizon.

The appearance of Sgr A* should reveal information about the orientation of the black hole's accretion disk along our line of sight and how fast the black hole is spinning—two of the most basic facts to be learned about the Sgr A* system and vital for understanding whatever else is observed about it (*below*). On occasions when a bright spot flares up in the accretion disk, gravitational lensing by the black hole will form multiple subimages of the spot (*opposite page*). If these subimages can be resolved, they will provide detailed information about the gravitational field near the black hole, which will stringently test the predictions of general relativity.



Collecting Data

The Combined Array for Research in Millimeter-Wave Astronomy (CARMA; *above*), located at Cedar Flat, Calif., is one of several radio telescope arrays astronomers have developed to observe Sgr A*'s event horizon. A network of such observatories (*left*) separated by baselines thousands of kilometers long (*lines*) can exploit a technique called very long baseline interferometry to produce images with resolutions as fine as those that would be possible with a radio dish the size of Earth. Four arrays (*green*) are ready to be used together, and three (*pink*) are under development to become new sites within the Event Horizon Telescope array.


Accretion disk
Event horizon
Simulation 1
Nonrotating black hole viewed from 30 degrees above accretion disk plane
Simulation 2
Nonrotating black hole viewed from 10 degrees above accretion disk plane
Simulation 3
Rapidly spinning black hole viewed from 10 degrees above accretion disk plane
55 microarcseconds

What the Silhouette Can Reveal

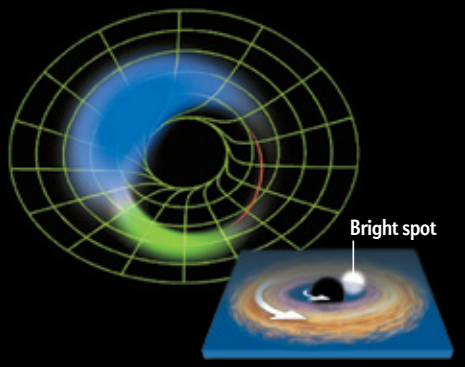
Simulations show how an accretion disk around Sgr A* would appear depending on the orientation of its accretion disk and the magnitude of its spin. The rightmost trio of images includes the blurring effects of interstellar gas.

The green coordinate grid is in the plane of the accretion disk, centered on the black hole. The grid's innermost ring is at the black hole's event horizon. Bending of light rays by the hole's gravity, known as gravitational lensing, distorts the grid's appearance and also magnifies the hole's silhouette. Because the accretion disk orbits the hole at velocities approaching the speed of light, special relativistic effects come into play, making it much brighter on the side moving toward us (*here on left side of event horizon*). In the bottom image, the black hole's large, angular momentum causes additional deflection of light, further distorting our view of the equatorial plane and dramatically changing the appearance of the accreting gas.

Thus, comparing images of Sgr A* with simulations can reveal the system's orientation and the black hole's spin and can also provide—from the silhouette's size—a new measurement of the hole's mass.

Illustration by Jen Christiansen

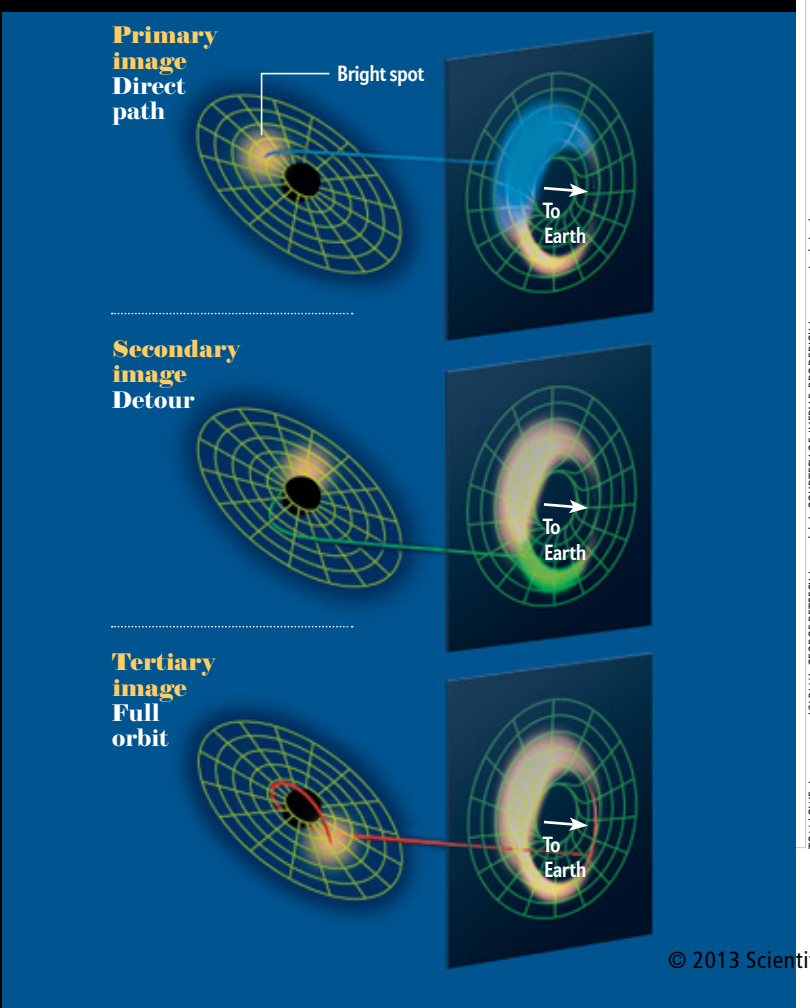
© 2013 Scientific American



Measuring Gravity with Lensed Images

Astronomers can measure gravity very near a black hole by analyzing multiple subimages (produced by gravitational lensing) of a bright spot in an accretion disk. Shown above is a simulated image of a bright spot near a moderately rotating black hole, colored to mark its three constituent subimages, which are explained in the diagrams below.

The primary image (*blue region*) forms from radio waves that traveled from the spot along the most direct path to Earth (*blue line*). Thanks to the hole's intense gravity, some rays the spot emitted earlier take a detour around the hole (*green line*) and reach Earth at the same time, forming the secondary image (*green region*). Rays that were emitted even earlier and execute a full orbit around the black hole (*red line*) generate the barely visible tertiary image (*red region*). Because the positions and shapes of the subimages depend on how gravity bends light in a variety of locations very close to the hole, analysis of the full image can reveal if general relativity correctly describes gravity there.



TOM LOWE *timescapes.org (CARMA)*; GEORGE RETSECK *(array globe)*; COURTESY OF AVERY E. BRODERICK *(computer simulations)*

first empirical evidence that jet production and black hole spin are related, something that has long been suspected on theoretical grounds.

For both Sgr A* and M87, an exciting prospect in the long run will be the possibility of imaging flare-ups that are seen in their emission from time to time. If some of these flares are caused by bright spots in the accretion flow, as most theorists expect, they can be exploited to map out the spacetime around the horizon in greater detail. The main image of each spot will be accompanied by additional images, corresponding to light rays that reach the observer by circuitous routes around the hole [*see box at left*]. The shapes and positions of these higher-order images encode the structure of the spacetime near the black hole. They will, in effect, provide independent measurements of that structure at the different locations traversed by each image's bundle of light rays. Taken together, these data will sternly test general relativity's predictions for the behavior of strong gravity near black holes.

Black hole observations are entering a new golden era. Almost a century after Einstein conceived of general relativity, we are finally in a position to test whether this theory correctly describes gravity in the extreme environments of black holes. Direct imaging of black holes will provide a new test bed for comparing Einstein's theory with its alternatives. When images of Sgr A* and M87 become available, we will be able to investigate the spacetime near black holes in detail, without sacrificing cell phone technicians. SA

Avery E. Broderick and **Abraham Loeb** began collaborating in 2005 at the Institute for Theory and Computation, which Loeb now directs, at the Harvard-Smithsonian Center for Astrophysics. Broderick is currently an assistant professor in the department of physics and astronomy at the University of Waterloo and an associate faculty member at the Perimeter Institute for Theoretical Astrophysics. He has been advancing the science of horizon-resolving imaging of supermassive black holes. Loeb is Frank B. Baird, Jr., Professor of Science at Harvard University and also serves as chair of the Harvard astronomy department. He pioneered theoretical studies of the first stars and black holes and recently wrote two books on the first galaxies in the universe.

MORE TO EXPLORE

Event-Horizon-Scale Structure in the Supermassive Black Hole Candidate at the Galactic Centre. Sheperd S. Doeleman et al. in *Nature*, Vol. 455, pages 78–80; September 4, 2008.

Imaging the Black Hole Silhouette of M87: Implications for Jet Formation and Black Hole Spin. Avery E. Broderick and Abraham Loeb in *Astrophysical Journal*, Vol. 697, pages 1164–1179; June 1, 2009.

The Event Horizon of Sagittarius A*. Avery E. Broderick, Abraham Loeb and Ramesh Narayan in *Astrophysical Journal*, Vol. 701, pages 1357–1366; August 20, 2009.

Evidence for Low Black Hole Spin and Physically Motivated Accretion Models from Millimeter-VLBI Observations of Sagittarius A*. Avery E. Broderick, Vincent L. Fish, Sheperd S. Doeleman and Abraham Loeb in *Astrophysical Journal*, Vol. 735, No. 2; July 10, 2011.

Jet-Launching Structure Resolved Near the Supermassive Black Hole in M87. Sheperd S. Doeleman et al. in *Science*, Vol. 338, pages 355–358; October 19, 2012.

Imaging an Event Horizon: Submm-VLBI of a Super Massive Black Hole. Sheperd S. Doeleman et al. in *ASTRO2010 Decadal Review*. http://arxiv.org/abs/0906.3899

Inside Black Holes. Andrew J. S. Hamilton. Includes animations of a descent into a black hole. http://jila.colorado.edu/~ajsh/insidebh

U.C.L.A. Galactic Center Group Web site: www.astro.ucla.edu/~ghezgroup/gc

© 2013 Scientific American

Clothed or Naked?

Black holes and naked singularities are two possible outcomes of the collapse of a dying massive star. At the heart of each is a singularity—a wad of matter so dense that it requires new laws of physics to describe. Anything that hits the singularity gets destroyed.

In a black hole, the singularity is "clothed"—that is, surrounded by a boundary called the event horizon that hides it. Nothing that falls through this surface can ever get back out. A naked singularity has no such boundary. It is visible to outside observers, and objects that fall toward the singularity can in principle reverse course right up to the moment of impact.

Observing Naked Singularities

Naked singularities could reveal their presence to astronomers in several ways:

- High-energy explosions that produce naked singularities would brighten and fade in a distinctive way. Certain unexplained classes of gamma-ray bursts might do this.
- The accretion disk around a naked singularity would have different properties than the accretion disk around a black hole and typically would be brighter.
- Naked singularities would bend the light of background stars differently than black holes do.
- Observations of mass-to-spin ratio could distinguish between a black hole and a faster-spinning naked singularity. The planned Square Kilometer Array (SKA) radio telescope would have the requisite precision to tell.

© 2013 Scientific American

GRAVITY

NAKED SINGULARITIES

The black hole has a troublesome sibling, the naked singularity. Physicists have long thought—hoped—it could never exist. But could it?

By Pankaj S. Joshi

MODERN SCIENCE HAS INTRODUCED THE WORLD TO PLENTY of strange ideas, but surely one of the strangest is the fate of a massive star that has reached the end of its life. Having exhausted the fuel that sustained it for millions of years, the star is no longer able to hold itself up under its own weight, and it starts collapsing catastrophically. Modest stars like the sun also collapse, but they stabilize again at a smaller size. Whereas if a star is massive enough, its gravity overwhelms all the forces that might halt the collapse. From a size of millions of kilometers across, the star crumples to a pinprick smaller than the dot on an "i."

Most physicists and astronomers think the result is a black hole, a body with such intense gravity that nothing can escape from its immediate vicinity. A black hole has two parts. At its core is a singularity, the infinitesimal point into which all the matter of the star gets crushed. Surrounding the singularity is the region of space from which escape is impossible, the perimeter of which is called the event horizon. Once something enters the event horizon, it loses all hope of exiting. Whatever light the falling body gives off is trapped, too, so an outside observer never sees it again. It ultimately crashes into the singularity.

But is this picture really true? The known laws of physics are clear that a singularity forms, but they are hazy about the event horizon. Most physicists operate under the assumption that a horizon must indeed form, if only because the horizon is very appealing as a scientific fig leaf. Physicists have yet to figure out what

IN BRIEF

Conventional wisdom has it that a large star eventually collapses to a black hole, but some theoretical models suggest it might instead become a so-called naked singularity [*see opposite page for definition*]. Sorting out what happens is one of the most important unresolved problems in astrophysics.

The discovery of naked singularities would transform the search for a unified theory of physics, not least by providing direct observational tests of such a theory.

Illustration by Kenn Brown, Mondolithic Studios

© 2013 Scientific American

exactly happens at a singularity: matter is crushed, but what becomes of it then? The event horizon, by hiding the singularity, isolates this gap in our knowledge. All kinds of processes unknown to science may occur at the singularity, yet they have no effect on the outside world. Astronomers plotting the orbits of planets and stars can safely ignore the uncertainties introduced by singularities and apply the standard laws of physics with confidence. Whatever happens in a black hole stays in a black hole.

Yet a growing body of research calls this working assumption into question. Researchers have found a wide variety of stellar collapse scenarios in which an event horizon does not in fact form, so that the singularity remains exposed to our view. Physicists call it a naked singularity. Matter and radiation can both fall in and come out. Whereas visiting the singularity inside a black hole would be a one-way trip, you could in principle come as close as you like to a naked singularity and return to tell the tale.

If naked singularities exist, the implications would be enormous and would touch on nearly every aspect of astrophysics and fundamental physics. The lack of horizons could mean that mysterious processes occurring near the singularities would impinge on the outside world. Naked singularities might account for unexplained high-energy phenomena that astronomers have seen, and they might offer a laboratory to explore the fabric of spacetime on its finest scales.

COSMIC CENSOR

EVENT HORIZONS were supposed to have been the easy part about black holes. Singularities are clearly mysterious. They are places where the strength of gravity becomes infinite and the known laws of physics break down. According to physicists' current understanding of gravity, encapsulated in Einstein's general theory of relativity, singularities inevitably arise during the collapse of a giant star. General relativity does not account for the quantum effects that become important for microscopic objects, and those effects presumably intervene to prevent the strength of gravity from becoming truly infinite. But physicists are still struggling to develop the quantum theory of gravity they need to explain singularities.

By comparison, what happens to the region of spacetime around the singularity seems as though it should be rather straightforward. Stellar event horizons are many kilometers in size, far larger than the typical scale of quantum effects. Assuming that no new forces of nature intervene, horizons should be governed purely by general relativity, a theory that is based on well-understood principles and has passed 90 years of observational tests.

That said, applying the theory to stellar collapse is still a formidable task. Einstein's equations of gravity are notoriously complex, and solving them requires physicists to make simplifying assumptions. American physicists J. Robert Oppenheimer and Hartland S. Snyder—and, independently, Indian physicist B. Datt—made an initial attempt in the late 1930s. To simplify the equations, they considered only perfectly spherical stars, assumed the stars consisted of gas of a homogeneous (uniform) density and neglected gas pressure. They found that as this idealized star collapses, the gravity at its surface intensifies and eventually becomes strong enough to trap all light and matter, thereby forming an event horizon. The star becomes invisible to outside observers and soon thereafter collapses all the way down to a singularity.

Real stars, of course, are more complicated. Their density is inhomogeneous, the gas in them exerts pressure, and they can assume other shapes. Does every sufficiently massive collapsing star turn into a black hole? In 1969 University of Oxford physicist Roger Penrose suggested that the answer is yes. He conjectured that the formation of a singularity during stellar collapse necessarily entails the formation of an event horizon. Nature thus forbids us from ever seeing a singularity because a horizon always cloaks it. Penrose's conjecture is termed the cosmic censorship hypothesis. It is only a conjecture, but it underpins the modern study of black holes. Physicists hoped we would be able to prove it with the same mathematical rigor we used to show the inevitability of singularities.

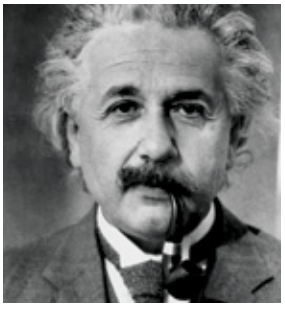
SINGULARITIES AU NATUREL

THAT HAS NOT HAPPENED. Instead of coming up with a direct proof of censorship that applies under all conditions, we have

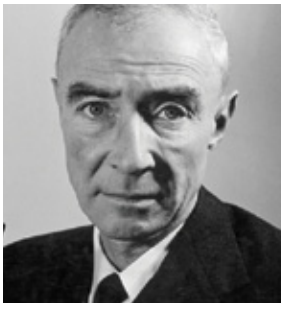
TRAILBLAZING

Father Figures

The ongoing debate over whether naked singularities can form is one part of a broader story arc about black holes.



The general theory of relativity predicted black holes, but Albert Einstein doubted they could ever really form.



J. Robert Oppenheimer (later head of the Manhattan Project) and others showed they could.

Stephen Hawking and Roger Penrose (*right*) proved that singularities were, in fact, unavoidable.

Penrose conjectured that singularities had to remain clothed by event horizons. Others disagree.

SCIENCE SOURCE (*Einstein*); CORBIS (*Oppenheimer*); MATT DUNHAM *Corbis* (*Hawking*); BOB MAHONEY *Getty Images* (*Penrose*)

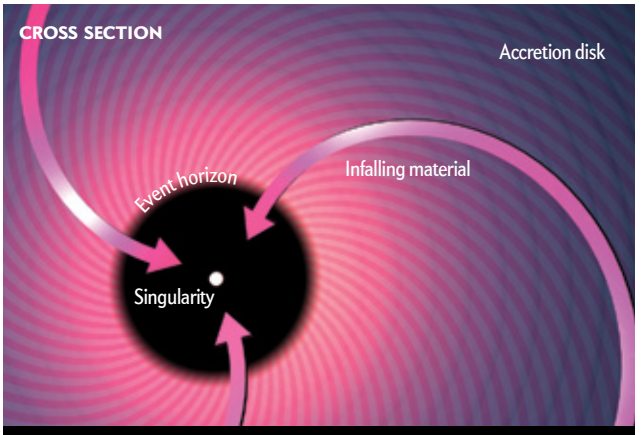
© 2013 Scientific American

BASIC PRINCIPLES

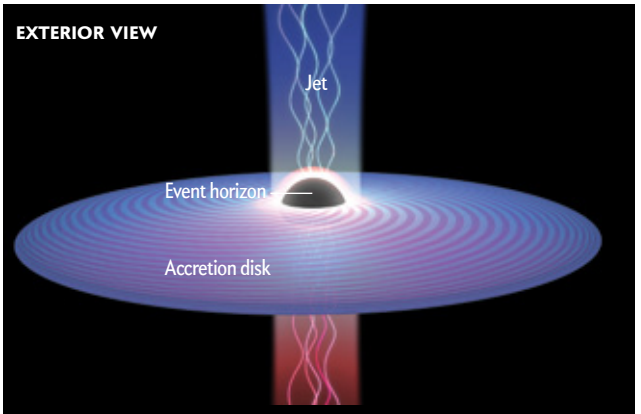
Two Monsters of the Cosmos

A naked singularity is essentially a black hole without the "black" part. It can both suck in and spit out matter and radiation. Accordingly, it would look different and have different effects on its immediate surroundings.

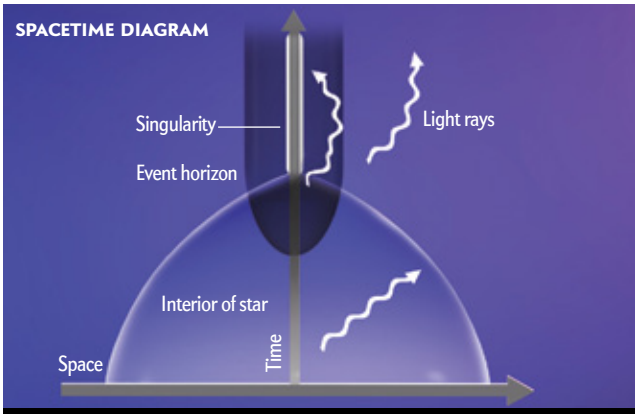
Black Hole



The key feature of a black hole is its event horizon, a surface that material can enter but not exit. Often the horizon is surrounded by a swirling, gaseous disk.

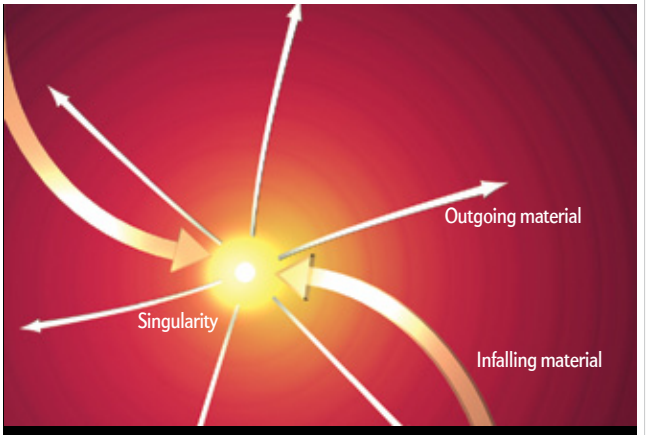


On the outside, the black hole looks like a pitch-black ball. The singularity lies within and cannot be seen. Friction within the surrounding disk generates intense radiation. Some disk material flies out in a jet; other material falls in.

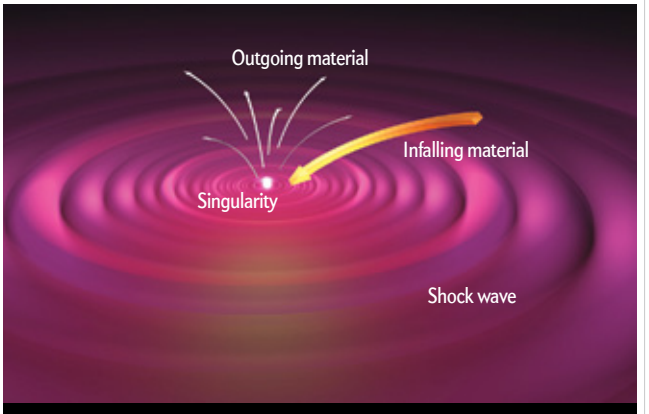


A homogeneous star without gas pressure collapses to a black hole. The star's gravity intensifies, bending the paths of moving objects (including light rays) by an increasing amount and eventually trapping them.

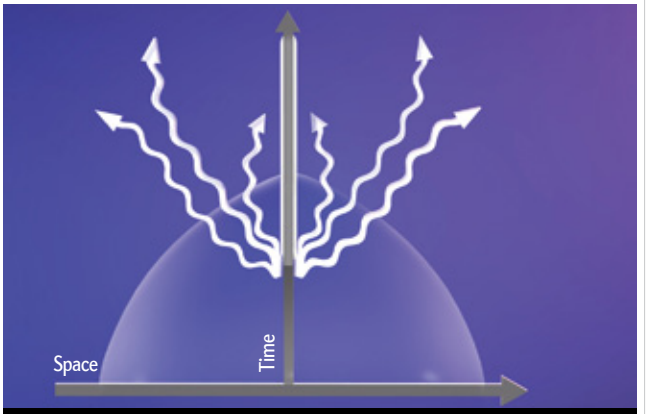
Naked Singularity



A naked singularity lacks an event horizon. Like a black hole, it can suck in material; unlike a black hole, it can also spit it out.



A naked singularity looks like a tiny dust grain, except that it is almost inconceivably dense. Infalling matter can be seen all the way down to its final impact with the singularity. The intense gravity can generate powerful shock waves.



When the star is inhomogeneous, its gravity might never intensify enough to bend light rays back on themselves. The star collapses to a singularity, but the singularity is visible. [*For step-by-step simulations, see box on next two pages.*]

Illustration by Alfred T. Kamajian

© 2013 Scientific American

had to embark on the longer route of analyzing case studies of gravitational collapse one by one, gradually embellishing our theoretical models with the features that the initial efforts lacked. In 1973 German physicist Hans Jürgen Seifert and his colleagues considered inhomogeneity. Intriguingly, they found that layers of infalling matter could intersect to create momentary singularities that were not covered by horizons. But singularities come in various types, and these ones were fairly benign. Although the density at one location became infinite, the strength of gravity did not, so the singularity did not crush matter and infalling objects to an infinitesimal pinprick. Thus, general relativity never broke down, and matter continued to move through this location rather than meeting its end.

In 1979 Douglas M. Eardley of the University of California, Santa Barbara, and Larry Smarr, then at the University of Illinois at Urbana-Champaign, went a step further and performed a numerical simulation of a star with a realistic density profile: highest at its center and slowly decreasing toward the surface. An exact paper-and-pencil treatment of the same situation, undertaken by Demetrios Christodoulou of the Swiss Federal Institute of Technology in Zurich, followed in 1984. Both studies found that the star shrank to zero size and that a naked singularity resulted. But the model still neglected pressure, and Richard P.A.C. Newman, then at the University of York in England, showed that the singularity was again gravitationally weak.

Inspired by these findings, many researchers, including me, tried to formulate a rigorous theorem that naked singularities would always be weak. We were unsuccessful. The reason soon became clear: naked singularities are not always weak. We found scenarios of inhomogeneous collapse that led to singularities where gravity was strong—that is, genuine singularities that could crush matter into oblivion—yet remained visible to external observers. A general analysis of stellar collapse in the absence of gas pressure, developed in 1993 by Indresh Dwivedi, then at Agra University, and me, clarified and settled these points.

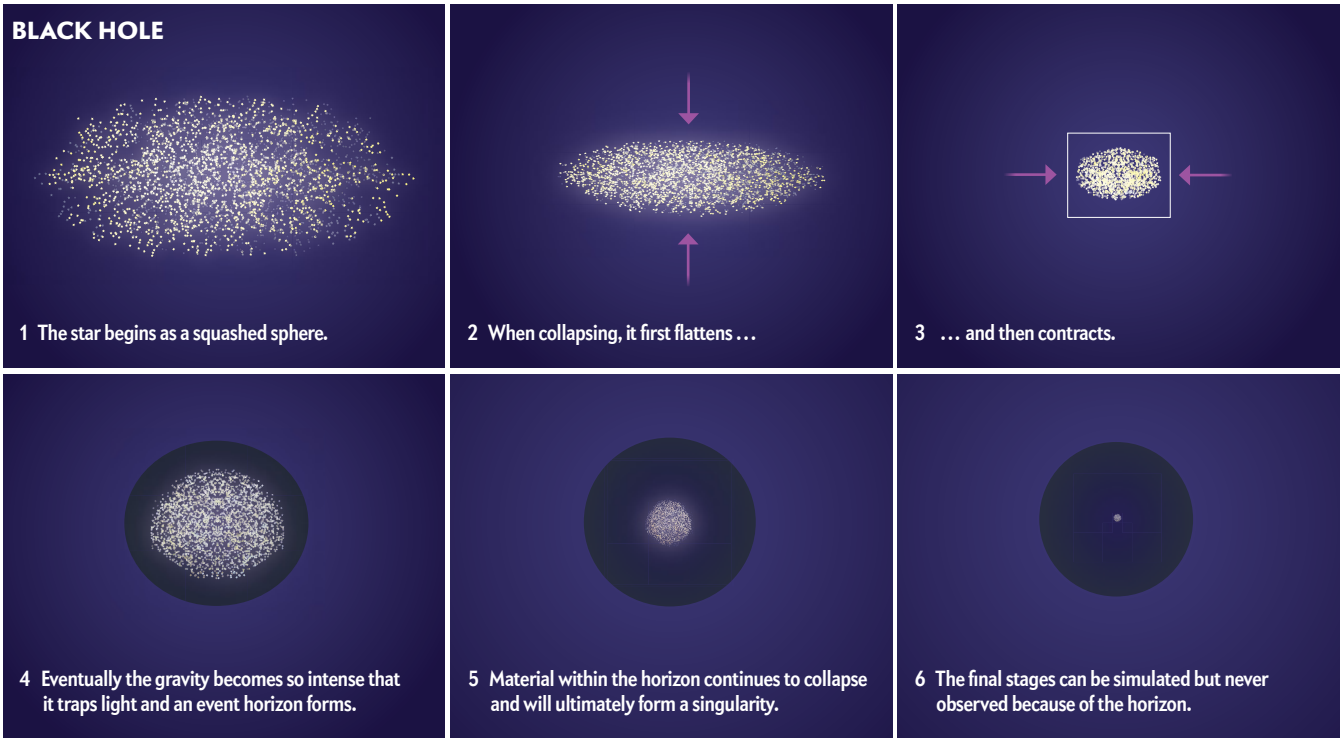
In the early 1990s physicists considered the effects of gas pressure. Amos Ori of the Technion-Israel Institute of Technology and Tsvi Piran of the Hebrew University of Jerusalem conducted numerical simulations, and my group solved the relevant equations exactly. Stars with a fully realistic relation between density and pressure could collapse to naked singularities. At about the same time, teams led by Giulio Magli of the Polytechnic University of Milan and by Kenichi Nakao of Osaka City University considered a form of pressure generated by rotation of particles within a collapsing star. They, too, showed that in a wide variety of situations, collapse ends in a naked singularity after all.

These studies analyzed perfectly spherical stars, which is not as severe a limitation as it might appear, because most stars in nature are very close to this shape. Moreover, spherical stars have, if anything, more favorable conditions for horizon formation than stars of other shapes do, so if cosmic censorship fails even for

COMPUTER SIMULATIONS

Two Ways to Crush a Star

Computer simulations reveal the circumstances under which a star collapses to a black hole or to a naked singularity. The simulations shown


BLACK HOLE
1 The star begins as a squashed sphere.
2 When collapsing, it first flattens . . .
3 . . . and then contracts.
4 Eventually the gravity becomes so intense that it traps light and an event horizon forms.
5 Material within the horizon continues to collapse and will ultimately form a singularity.
6 The final stages can be simulated but never observed because of the horizon.

SOURCE: STUART L. SHAPIRO *University of Illinois* AND SAUL A. TEUKOLSKY *Cornell University (computer simulations)*

Illustration by Lila Rubenstein

© 2013 Scientific American

them, its prospects look questionable. That said, physicists have been exploring nonspherical collapse. In 1991 Stuart L. Shapiro of the University of Illinois and Saul A. Teukolsky of Cornell University presented numerical simulations in which oblong stars could collapse to a naked singularity. A few years later Andrzej Królak of the Polish Academy of Sciences and I studied nonspherical collapse and also found naked singularities. To be sure, both these studies neglected gas pressure.

Skeptics have wondered whether these situations are contrived. Would a slight change to the initial configuration of the star abruptly cause an event horizon to cover the singularity? If so, then the naked singularity might be an artifact of the approximations used in the calculations and would not truly arise in nature. Some scenarios involving unusual forms of matter are indeed very sensitive. But our results so far also show that most naked singularities are stable to small variations of the initial setup. Thus, these situations appear to be what physicists call generic—that is, they are not contrived.

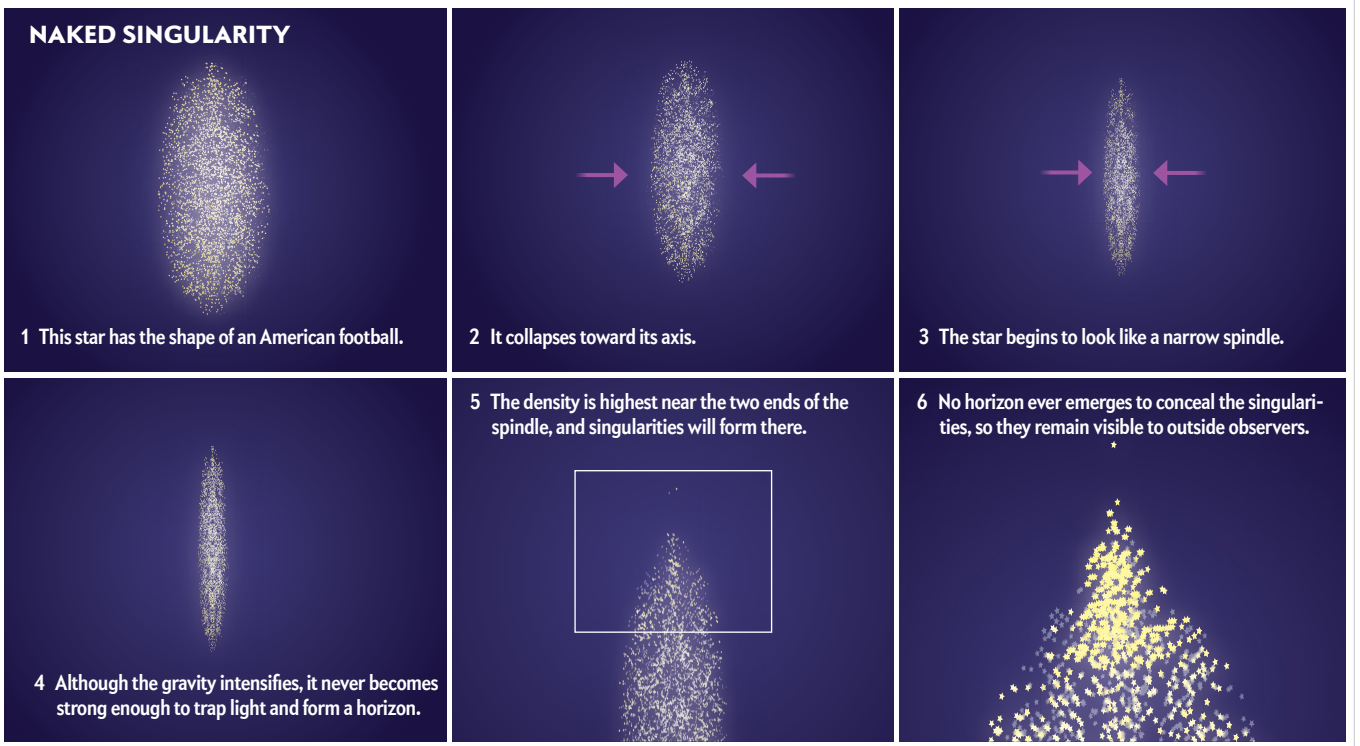
HOW TO BEAT THE CENSOR

THESE COUNTEREXAMPLES to Penrose's conjecture suggest that cosmic censorship is not a general rule. Physicists cannot say, "Collapse of any massive star makes a black hole only," or "Any physically realistic collapse ends in a black hole." Some scenarios lead to a black hole and others to a naked singularity. In some models, the singularity is visible only temporarily, and an event horizon eventually forms to cloak it. In others, the singularity remains visible forever. Typically the naked singularity develops in the geometric center of collapse, but it does not always do so, and even when it does, it can also spread to other regions. Nakedness also comes in degrees: an event horizon might hide the singularity from the prying eyes of faraway observers, whereas observers who fell through the event horizon could see the singularity prior to hitting it. The variety of outcomes is bewildering.

My colleagues and I have isolated various features of these scenarios that cause an event horizon to arise or not. In particular, we have examined the role of inhomogeneities and gas pressure. According to Einstein's theory, gravity is a complex phenomenon involving not only a force of attraction but also effects such as shearing, in which different layers of material are shifted laterally in opposite directions. If the density of a collapsing star is very high—so high that by all rights it should trap light—but also inhomogeneous, those other effects may create escape routes. Shearing of material close to a singularity, for example, can set off powerful shock waves that eject matter and light—in essence, a gravitational typhoon that disrupts the formation of an event horizon.

To be specific, consider a homogeneous star, neglecting gas pressure. (Pressure alters the details but not the broad outlines of what happens.) As the star collapses, gravity increases in strength and bends the paths of moving objects ever more severely. Light rays, too, become bent, and there comes a time when the bending is so severe that light can no longer propagate away from the star. The region where light becomes trapped starts off small, grows and eventually reaches a stable

here treat the star as a swarm of grains whose gravity is so powerful that other forces of nature, such as gas pressure, are unimportant.


NAKED SINGULARITY
1 This star has the shape of an American football.
2 It collapses toward its axis.
3 The star begins to look like a narrow spindle.
4 Although the gravity intensifies, it never becomes strong enough to trap light and form a horizon.
5 The density is highest near the two ends of the spindle, and singularities will form there.
6 No horizon ever emerges to conceal the singularities, so they remain visible to outside observers.

© 2013 Scientific American

size proportional to the star's mass. Meanwhile because the star's density is uniform in space and varies only in time, the entire star is crushed to a point simultaneously. The trapping of light occurs well before this moment, so the singularity remains hidden.

Now consider the same situation except that the density decreases with distance from the center. In effect, the star has an onionlike structure of concentric shells of matter. The strength of gravity acting on each shell depends on the average density of matter within that shell. Because the denser inner shells feel a stronger pull, they collapse faster than the outer ones. The entire star does not collapse to a singularity simultaneously. The innermost shells collapse first, and then the outer shells pile on, one by one.

The resulting delay can postpone the formation of an event horizon. If the horizon can form anywhere, it will form in the dense inner shells. But if density decreases with distance too rapidly, these shells may not constitute enough mass to trap light. The singularity, when it forms, will be naked. Thus, there is a threshold: if the degree of inhomogeneity is very small, below a critical limit, a black hole will form; with sufficient inhomogeneity, a naked singularity arises.

In other scenarios, the salient issue is the rapidity of collapse. This effect comes out very clearly in models where stellar gas has converted fully to radiation and, in effect, the star becomes a giant fireball—a scenario first considered by the late Indian physicist P. C. Vaidya in the 1940s. Again there is a threshold: slowly collapsing fireballs become black holes, but if a fireball collapses rapidly enough, light does not become trapped.

SLIME AND SOCKS

ONE REASON it has taken so long for physicists to accept the possibility of naked singularities is that they raise a number of conceptual puzzles. A commonly cited concern is that such singularities would make nature inherently unpredictable. Because general relativity breaks down at singularities, it cannot predict what those singularities will do. John Earman of the University of Pittsburgh memorably suggested that green slime and lost socks could emerge from them.

As long as singularities remain safely ensconced within event horizons, this randomness remains contained and general relativity is a fully predictive theory, at least outside the horizon. But if singularities can be naked, their unpredictability would infect the rest of the universe. For example, when physicists applied general relativity to Earth's orbit around the sun, they would in effect have to make allowance for the possibility that a singularity somewhere in the universe could emit a random gravitational pulse and send our planet flying off into deep space.

Yet this worry is misplaced. Unpredictability is actually common in general relativity—and not always directly related to censorship violation. The theory permits time travel, which could produce causal loops with unforeseeable outcomes, and even ordinary black holes can become unpredictable. For example, if we drop an electrical charge into an uncharged black hole, the shape of spacetime around the hole radically changes and is no longer predictable. A similar situation holds when the black hole is rotating. Specifically, what happens is that spacetime no longer neatly separates into space and time, so physicists cannot consider how the black hole evolves from some initial time into the future. Only the purest of pure black holes, with no charge or rotation at all, is fully predictable.

The loss of predictability and other problems with black holes actually stem from the occurrence of singularities; it does not matter whether they are hidden or not. The solution probably lies in a quantum theory of gravity, which will go beyond general relativity and offer a full explication of singularities. Within that theory, every singularity would prove to have a high but finite density. A naked singularity would be a "quantum star," a hyperdense body governed by the rules of quantum gravity.

Another possibility is that singularities may really have an infinite density after all—that they are not things to be explained away by quantum gravity but to be accepted as they are. The breakdown of general relativity at such a location may not be a failure of the theory per se but a sign that space and time have an edge. The singularity marks the place where the physical world ends. We should think of it as an event rather than an object, a moment when collapsing matter reaches the edge and ceases to be, like the big bang in reverse.

In that case, questions such as what will come out of a naked singularity are

A RELATED PROBLEM

Can a Black Hole Be Cracked Open?

Besides the collapse of a star, another way to create a naked singularity might be to take an existing black hole and destroy it. Although that sounds like an impossible, not to mention dangerous, task, the equations of general relativity show that an event horizon can exist only if the hole does not rotate too fast or does not have too much electrical charge. Most physicists think that the hole resists any efforts to spin or charge it up beyond the prescribed limits. But some think the hole might eventually succumb, causing the horizon to dissipate and expose the singularity.

Spinning up a black hole is not terribly difficult. Matter naturally falls in with angular momentum, which drives the hole to spin faster, like pushing on a revolving door. Charging up a black hole is harder because a charged hole repels particles of the same charge and draws in oppositely charged ones, thereby neutralizing itself. But a vigorous infall of matter could overcome this tendency. The fundamental characteristic of a black hole—namely, its trait of gobbling up the matter around it that keeps it growing—could become the cause of its own destruction. Researchers continue to debate whether the black hole would eventually save itself or would crack open to reveal the singularity within. *—P.S.J.*

© 2013 Scientific American

Varieties of Stellar Collapse

Like humans, stars have a life cycle. They are born in gigantic clouds of dust and galactic material in the depths of space, evolve and shine for millions of years, and eventually enter a phase of dissolution and extinction. Stars shine by burning their nuclear fuel, which is initially mainly hydrogen; they fuse it into helium and, later, heavier elements. Each star attains a balance between the force of gravity, which pulls matter toward the center, and the outward pressures generated by fusion. This balance keeps the star stable—until all the fuel is converted to iron, which, in nuclear terms, is inert. Fusion then ceases, the all-pervasive force of gravity asserts itself, and the star begins to contract.

When our sun runs out of fuel, its core will contract under its own gravity until it is no bigger than Earth, at which point it will be supported by the force exerted by fast-moving electrons, called electron degeneracy pressure. The resulting object is called a white dwarf. Stars three to five times the mass of the sun settle to a different final state, a neutron star, in which gravity is so strong that atoms buckle, too. Supported by the pressure not of electrons but of neutrons, a neutron star is barely 10 kilometers in size.

Still more massive stars cannot settle either to a white dwarf or to a neutron star because these forms of pressure are just not sufficient. Unless some other, unknown form of pressure comes into play, gravitational collapse becomes unstoppable. Gravity now is the sole operative force, and the final fate of the star is determined by Einstein's theory of gravitation. The theory indicates that the outcome is a singularity, and the question is whether this singularity is visible or not. —*P.S.J.*

not really meaningful; there is nothing to come out of, because the singularity is just a moment in time. What we see from a distance is not the singularity itself but the processes that occur in the extreme conditions of matter near this event, such as shock waves caused by inhomogeneities in this ultradense medium.

In addition to unpredictability, a second issue troubles many physicists. Having provisionally assumed that the censorship conjecture holds, they have spent the past several decades formulating laws that black holes should obey. But the laws are not free of major paradoxes. For example, they hold that a black hole swallows and destroys information—which appears to contradict the basic principles of quantum theory. This paradox and other predicaments stem from the presence of an event horizon. If the horizon goes away, these problems might go away, too. For instance, if the star could radiate away most of its mass in the late stages of collapse, it would destroy no information and leave behind no singularity. In that case, it would not take a quantum theory of gravity to explain singularities; general relativity might do the trick itself.

A LAB FOR QUANTUM GRAVITY

FAR FROM CONSIDERING naked singularities a problem, physicists can see them as an asset. If the singularities that formed in the gravitational collapse of a massive star are visible to external observers, they could provide a laboratory to study quantum-gravitational effects. Quantum gravity theories in the making, such as string theory and loop quantum gravity, are badly in need of some kind of observational input, without which it is nearly impossible to constrain the plethora of possibilities. Physicists commonly seek that input in the early universe, when conditions were so extreme that quantum-gravitational effects dominated. But the big bang was a unique event. If singularities could be naked, they would allow astronomers to observe the equivalent of a big bang every time a massive star in the universe ends its life.

To explore how naked singularities might provide a glimpse into otherwise unobservable phenomena, we recently simulated how a star collapses to a naked singularity, taking into account the effects predicted by loop quantum gravity. According to this theory, space consists of tiny atoms, which become conspicuous when matter becomes sufficiently dense; the result is an extremely powerful repulsive force that prevents the density from ever becoming infinite. In our model, such a repulsive force dispersed the star and dissolved the singularity. Nearly a quarter of the mass of the star was ejected within the final fraction of a microsecond. Just before it did so, a faraway observer would have seen a sudden dip in the intensity of radiation from the collapsing star—a direct result of quantum-gravitational effects.

The explosion would have unleashed high-energy gamma rays, cosmic rays and other particles such as neutrinos. Upcoming experiments such as the Extreme Universe Space Observatory—hosted by the Japanese Experiment Module of the International Space Station and expected to fly in 2017—may have the needed sensitivity to see this emission. Because the details of the outpouring depend on the specifics of the quantum gravity theory, observations would provide a way to discriminate among theories.

Either proving or disproving cosmic censorship would create a mini explosion of its own within physics because naked singularities touch on so many deep aspects of current theories. What comes out unambiguously from the theoretical work so far is that singularities can be naked if the conditions are suitable. The question remains whether these conditions could ever arise in nature. SA

Pankaj S. Joshi is a physics professor at the Tata Institute of Fundamental Research in Mumbai. He specializes in gravitation and cosmology. His other interests include Indian and Western classical music and exploring the scientific content of ancient wisdom.

MORE TO EXPLORE

Black Holes, Naked Singularities and Cosmic Censorship. Stuart L. Shapiro and Saul A. Teukolsky in *American Scientist*, Vol. 79, No. 4, pages 330–343; July/August 1991.

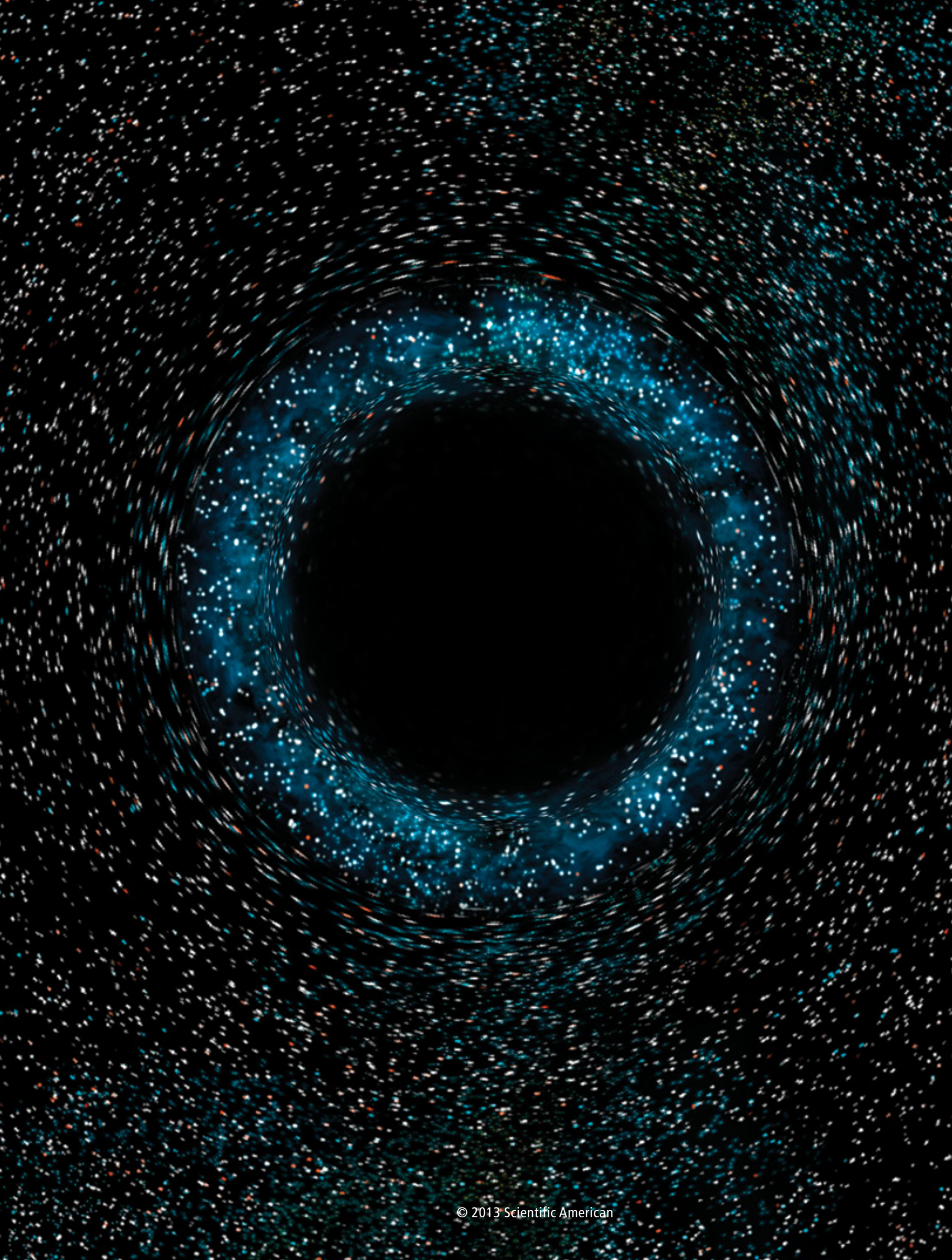
Black Holes and Time Warps. Kip S. Thorne. W. W. Norton, 1994.

Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes. John Earman. Oxford University Press, 1995.

Quantum Evaporation of a Naked Singularity. Rituparno Goswami, Pankaj S. Joshi and Parampreet Singh in *Physical Review Letters*, Vol. 96, No. 3, Article No. 031302; January 27, 2006. http://arxiv.org/abs/gr-qc/0506129

Gravitational Collapse and Space-time Singularities. Pankaj S. Joshi. Cambridge University Press, 2007.

© 2013 Scientific American



© 2013 Scientific American

ASTROPHYSICS

BLACK STARS, NOT HOLES

Quantum effects may prevent true black holes from forming and give rise instead to dense entities called black stars

By Carlos Barceló, Stefano Liberati, Sebastiano Sonego and Matt Visser

BLACK HOLES HAVE BEEN A PART OF POPULAR CULTURE FOR DECADES now. They played a central role in the plot of the 2009 *Star Trek* movie, for example. No wonder. These dark remnants of collapsed stars seem almost designed to bring out some of our primal fears: a black hole harbors unfathomable mystery behind the curtain that is its "event horizon," admits of no escape for anyone or anything that falls within, and irretrievably destroys all it ingests.

To theoretical physicists, black holes are a class of solutions of the Einstein field equations, which are at the heart of Einstein's general theory of relativity. The theory describes how all matter and energy distort spacetime as if it were made of elastic and how the resulting curvature of spacetime controls the motion of the matter and energy, producing the force we know as gravity. These equations unambiguously predict that there can be regions of spacetime from which no signal can reach distant observers. These regions—black holes—consist of a location where matter densities approach infinity (a "singularity") surrounded by an empty zone of extreme gravitation from which nothing, not even light, can escape. A conceptual boundary, the event horizon, separates the zone of intense gravitation from the rest of spacetime. In the simplest case, the event horizon is a sphere—just six kilometers in diameter for a black hole of the sun's mass.

So much for fiction and theory. What about reality? A wide variety of high-quality astrophysical observations indicates that the universe does contain some extremely compact bodies that emit essentially no light or other radiation of their own. Although these dark objects have masses ranging from just a few suns to well over a million suns, their

IN BRIEF

Black holes are theoretical structures in spacetime predicted by Einstein's general theory of relativity. Nothing can escape a black hole's gravity after passing inside its event horizon.

Approximate quantum calculations predict that black holes slowly evaporate, albeit in a paradoxical way. Physicists are still seeking a full, consistent quantum theory of gravity to describe black holes.

Contrary to physicists' conventional wisdom, a quantum effect called vacuum polarization may grow large enough to stop a hole from forming and create a "black star" instead.

COURTESY OF EUROPEAN SPACE AGENCY, NASA AND FELIX MIRABEL *French Atomic Energy Commission, Institute for Astronomy and Space Physics/Conicet of Argentina*

© 2013 Scientific American

diameters, as best astrophysicists can determine, range from only several kilometers to millions of kilometers—matching general relativity's predictions for black holes of those masses.

Yet are these dark and compact bodies that astronomers observe really the black holes predicted by general relativity? The observations to date certainly fit the theory quite well, but the theory itself is not entirely satisfactory in the way that it describes black holes. In particular, general relativity's prediction that a singularity resides inside every black hole suggests that the theory fails at that location, as is usually the case when a theory predicts that some quantity is infinite. Presumably general relativity fails by not taking into account quantum effects, which matter and energy exhibit at the microscopic scale. The search for a modified theory that incorporates quantum mechanics, generically called quantum gravity, is a powerful engine driving a lot of activity in theoretical physics research.

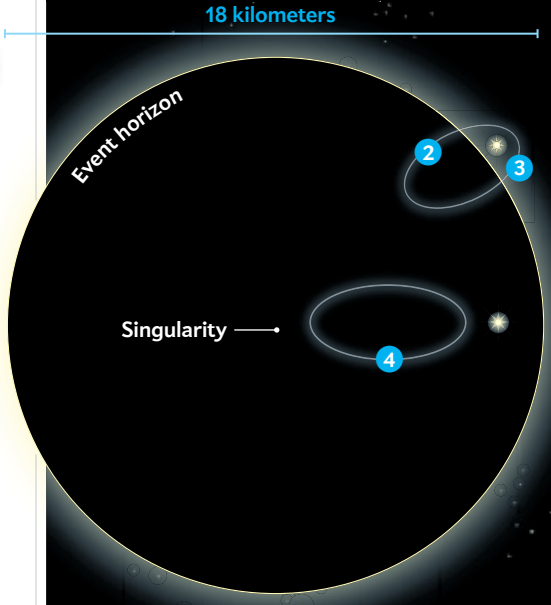
This need for a quantum theory of gravity raises fascinating questions: What would quantum-corrected black holes be like? Would they be radically different from classical black holes, or would their classical description remain a good approximation? The four of us have shown that certain quantum effects may well prevent black holes from forming at all. Instead a kind of object we have named a black star could arise. A black star would be blocked from taking the final plunge to infinite density and from becoming enveloped in an event horizon. The black star would be supported by something not normally considered to be a sturdy construction material: space itself.

BASICS

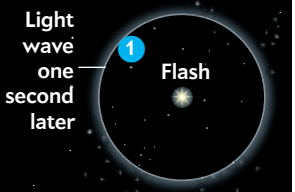
Black Holes in Brief

A black hole is a region of curved spacetime with such intense gravity that nothing can escape. Its defining feature is its event horizon: the boundary of the region of no escape. A black hole is mostly empty, its mass apparently collapsed to a location with infinite density—a "singularity"—deep inside the horizon.

A black hole with three times the mass of the sun would have a diameter of about 18 kilometers, comparable to the length of Manhattan.



Far away from large masses, a flash of light spreads out symmetrically in all directions 1.

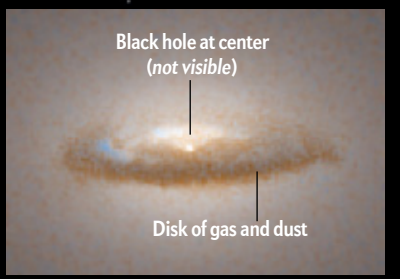


Just outside a black hole's event horizon, the gravity captures most of a flash 2.

Some light escapes, just 3.

If a flash occurs anywhere inside an event horizon, all the light is drawn into the black hole's singularity 4.

In practice, black holes can be observed via the material orbiting and falling into them. The image at the right, taken in 1998 by the Hubble Space Telescope, shows a vast disk of gas and dust believed to have a supermassive black hole at its center. Strictly speaking, however, such observations inform scientists only that an extremely compact, heavy object emitting little or no light of its own is present; they do not provide absolute proof that the object is a black hole.



COURTESY OF NASA, ROELAND P. VAN DER MAREL *Space Telescope Institute* AND FRANK C. VAN DEN BOSCH *University of Washington (dust disk around black hole)*

THE GRAVITY OF "EMPTY" SPACE

WE DERIVE OUR CONCLUSIONS by applying a venerable approach known as semiclassical gravity, but without making all the same assumptions about the collapsing matter that previous studies have made—to see if we might avoid the paradoxical territory arrived at by those studies. In the absence of a full-fledged theory of quantum gravity, theorists have resorted to semiclassical gravity over the past 30-odd years to analyze how quantum mechanics alters black holes. This method partially incorporates aspects of quantum physics—in particular, quantum field theory—into classical Einsteinian gravity.

Quantum field theory describes each kind of fundamental particle—the electron, the photon, quarks, you name it—in terms of a field that fills space, much like the electromagnetic field. Quantum field theory's equations are usually set up in flat spacetime, that is, in the absence of gravity. Semiclassical gravity uses quantum field theory as formulated in curved spacetime.

In the broadest terms, the strategy of semiclassical gravity goes as follows: a collection of matter in some configuration would, according to classical general relativity, produce some specific curved spacetime. Yet the curvature of spacetime modifies the energy of the quantum fields. This modified energy, according to classical general relativity, changes spacetime's curvature. And so on, iteration after iteration.

The goal is to obtain a self-consistent solution—a curved spacetime containing a configuration of quantum fields whose energy generates that same curvature. That kind of self-consistent solution ought to be a good approximation

Illustrations by Lucy Reading-Ikkanda

© 2013 Scientific American

to how reality behaves in many situations involving quantum effects and gravity even though gravity itself has not been described by a quantum theory. Semiclassical gravity thus incorporates quantum corrections into general relativity in a "minimal" way, taking into account the quantum behavior of matter but still treating gravity (that is, spacetime curvature) classically.

This approach, however, immediately runs into an embarrassing problem in that the straightforward calculation of the quantum fields' lowest possible (or "zero point") energy—the energy when no particles of any kind are present, the energy of the vacuum—produces an infinite result. This problem actually comes up already with ordinary quantum field theory (that is, in flat space, no gravity). Fortunately for theorists wishing to predict particle physics phenomena that do not involve gravity, the particles behave in ways that depend on only the energy differences between states, so that the value of the quantum vacuum energy plays no role. Careful subtraction schemes known as renormalization take care of the infinities, allowing the energy differences to be computed with extremely high precision.

With gravity in the picture, however, the vacuum energy matters. An infinite energy density would seem to produce an extremely large curvature of spacetime—that is, even "empty" space would harbor an intense gravitational force, which is not remotely compatible with the universe that we actually observe. Astronomical observations over the past decade indicate that the net zero-point contribution to the universe's total energy density is extremely tiny. The semiclassical gravity approach does not attempt to solve this problem. Instead it is customary to assume that whatever the solution is, it exactly cancels the zero-point contribution to the energy density in flat spacetime. This assumption makes for a consistent semiclassical vacuum: the energy density is zero everywhere, for which general relativity predicts flat spacetime.

If some matter is present, spacetime is curved, which alters the quantum fields' zero-point energy density, which means the zero-point energy is no longer exactly canceled. The excess amount is said to be caused by vacuum polarization, by analogy with the effect of an

Black Hole Categories

General relativity predicts that a black hole is completely defined by just three quantities: mass, angular momentum and electric charge. It makes no difference what went into the hole—matter, antimatter or energy, or all three combined.

Astronomers have observed holes in three mass classes: Holes of about five to 15 solar masses are formed from dying stars. Many galaxies harbor a hole of millions to billions of solar masses at their core. Holes of a few thousand solar masses have been detected in the center of globular star clusters.

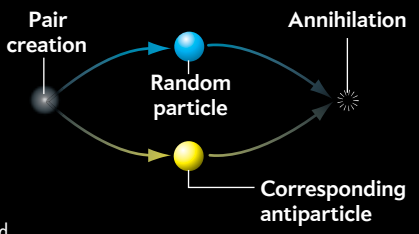
PARADOX

The Trouble with Quantum Black Holes

The classical (that is, nonquantum) equations of general relativity forbid anything emerging from inside a black hole's event horizon. Yet in the 1970s Stephen Hawking carried out quantum calculations that predicted black holes would randomly emit particles at a very low rate (*left panel*). The randomness created a paradoxical scenario (*right panel*) known as the information problem.

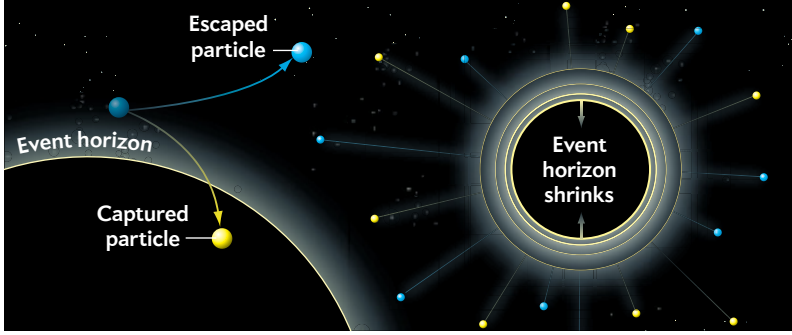
HAWKING RADIATION IS EMITTED

Even in empty space, a quantum process constantly produces pairs of so-called virtual particles and antiparticles, which immediately annihilate each other.



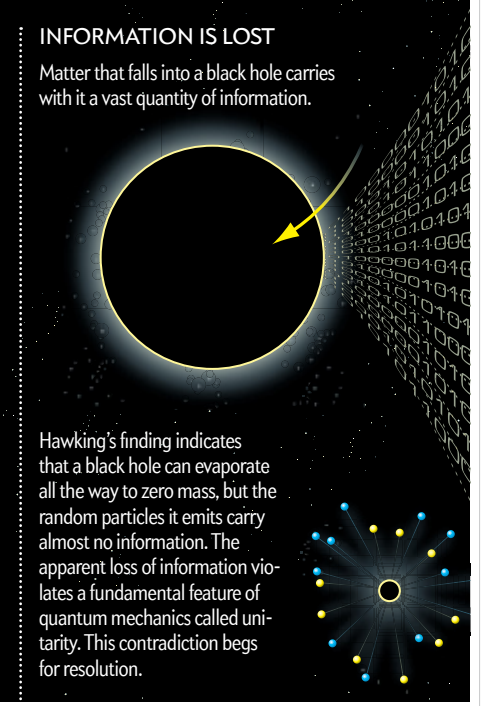
Near a black hole's event horizon, one virtual particle may be captured by the black hole, and the second may escape. The escaped particle carries away positive mass, and the captured one takes negative mass into the black hole—thereby reducing the hole's mass.

Thus, if nothing falls into the black hole, its mass and its event horizon gradually shrink. This evaporation process speeds up as the hole becomes smaller.



INFORMATION IS LOST

Matter that falls into a black hole carries with it a vast quantity of information.



Hawking's finding indicates that a black hole can evaporate all the way to zero mass, but the random particles it emits carry almost no information. The apparent loss of information violates a fundamental feature of quantum mechanics called unitarity. This contradiction begs for resolution.

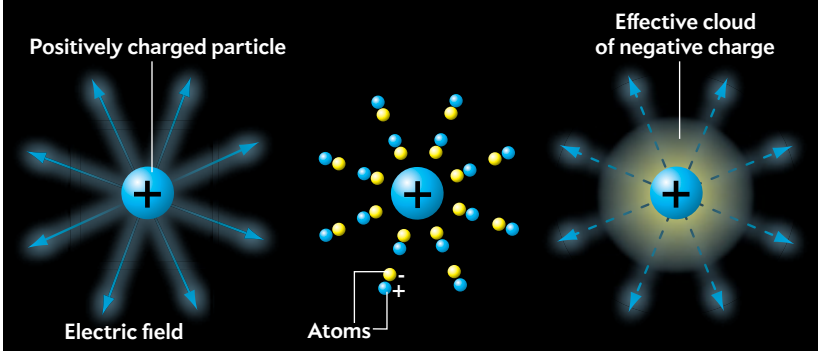
© 2013 Scientific American

What Emptiness Can Do

In classical general relativity, spacetime is dynamic, its curvature producing gravity. A quantum effect known as vacuum polarization provides another way that empty space can play an active role in the universe.

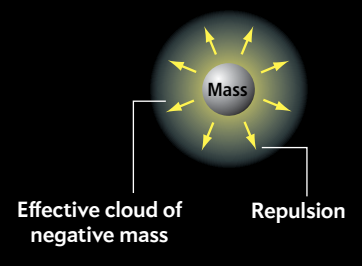
ELECTRIC ANALOGY

In a medium, a charged object's electric field (*left*) polarizes nearby atoms (*center*), reducing the total electric field (*right*). Quantum field theory reveals that even a vacuum can be polarized because an electric field polarizes virtual particle/antiparticle pairs.



VACUUM POLARIZATION

In general relativity, the role of electric charge is played by mass and energy and that of the electric field by curved spacetime, or gravity. The vacuum polarization produces an energy deficit (in effect a cloud of negative energy) and a repulsive force.



electric charge polarizing a medium [*see box above*].

We have described these features of semiclassical gravity in terms of mass and energy density, but in general relativity it is not only those quantities that produce spacetime curvature. The momentum density and the pressures and stresses associated with a specific gravitating substance also do so. A single mathematical-physics object, known as the stress energy tensor (SET), describes all these curvature-producing quantities. Semiclassical gravity assumes that the quantum fields' zero-point contributions to the total SET are exactly canceled in flat spacetime. The mathematical-physics object obtained applying such a subtraction procedure to the SET is called the renormalized stress energy tensor (RSET).

When applied in curved spacetime, the subtraction scheme still succeeds in canceling the SET's divergent part but leaves a finite, nonzero value for the RSET. The end result is an iterative process: classical matter curves spacetime via Einstein's equations, by an amount determined by the matter's classical SET. This curvature makes the quantum vacuum acquire a finite nonzero RSET. This vacuum RSET becomes an additional source of gravity, modifying the curvature. The new curvature induces in turn a different RSET, and so on.

QUANTUM-CORRECTED HOLES

WITH THE APPROACH of semiclassical gravity spelled out, the question becomes: How do these quantum corrections affect predictions about black holes? In particular, how do the corrections alter the process of forming a black hole?

The simplest black hole of some mass (say, M times the solar mass) is one that is not rotating and not electrically charged. Such a hole has a radius R that works out to be $3M$ kilometers. The radius R is called the gravitational, or Schwarzschild, radius, for that mass. If for any reason some matter has collapsed to occupy a region smaller than its gravitational radius, it has formed a black hole; it disappears inside its own event horizon.

The sun, for instance, has a 700,000-kilometer radius, which is much larger than its gravitational radius (three kilometers). The relevant semiclassical gravity equations make it clear that the RSET of the quantum vacuum in this situation is negligible. Thus, the sun is far from forming a black hole according to the classical equations, and quantum corrections do not alter this picture. Indeed, astrophysicists can safely ignore quantum gravity effects when analyzing the sun and most other astronomical objects.

The quantum corrections can become significant, however, if a star is not much larger than its gravitational radius. In 1976 David G. Boulware, now at the University of Washington, analyzed the case of such a compact star when the star is stationary (that is, not collapsing). He showed that the closer the star is to its gravitational radius, the larger the vacuum RSET near its surface becomes, increasing to infinite energy density. This result implies that semiclassical gravity theory does not permit a stationary black hole (meaning one whose event horizon remains constant in size) as a solution of its equations.

Boulware's result, however, does not tell us what to expect in the case of a star undergoing a collapse that would lead to a black hole according to classical general relativity. Stephen Hawking had already tackled this situation a year earlier, using somewhat different techniques, to show that a classical black hole formed by collapse emits random particles. More precisely, the particles have a distribution of energies characteristic of thermal radiation; the black hole has a temperature. He conjectured that quantum-corrected black holes would be essentially classical black holes subject to slow evaporation via this radiation. A black hole of one solar mass has a temperature of 60 nanokelvins. The corresponding evaporation rate is so slow that absorption of cosmic background radiation would completely overwhelm the evaporation and the hole would grow in size. An evaporating black hole

© 2013 Scientific American

of such a mass would be indistinguishable from a classical black hole in practice because the evaporation would be immeasurably small.

Considerable effort by theorists in the decade after Hawking's paper, including the approximate calculation of the RSET in collapsing configurations, reinforced this picture as being the correct one. Today the standard view in the physics community is that black holes form as described by classical general relativity and subsequently undergo slow quantum evaporation via Hawking radiation.

THE INFORMATION PROBLEM

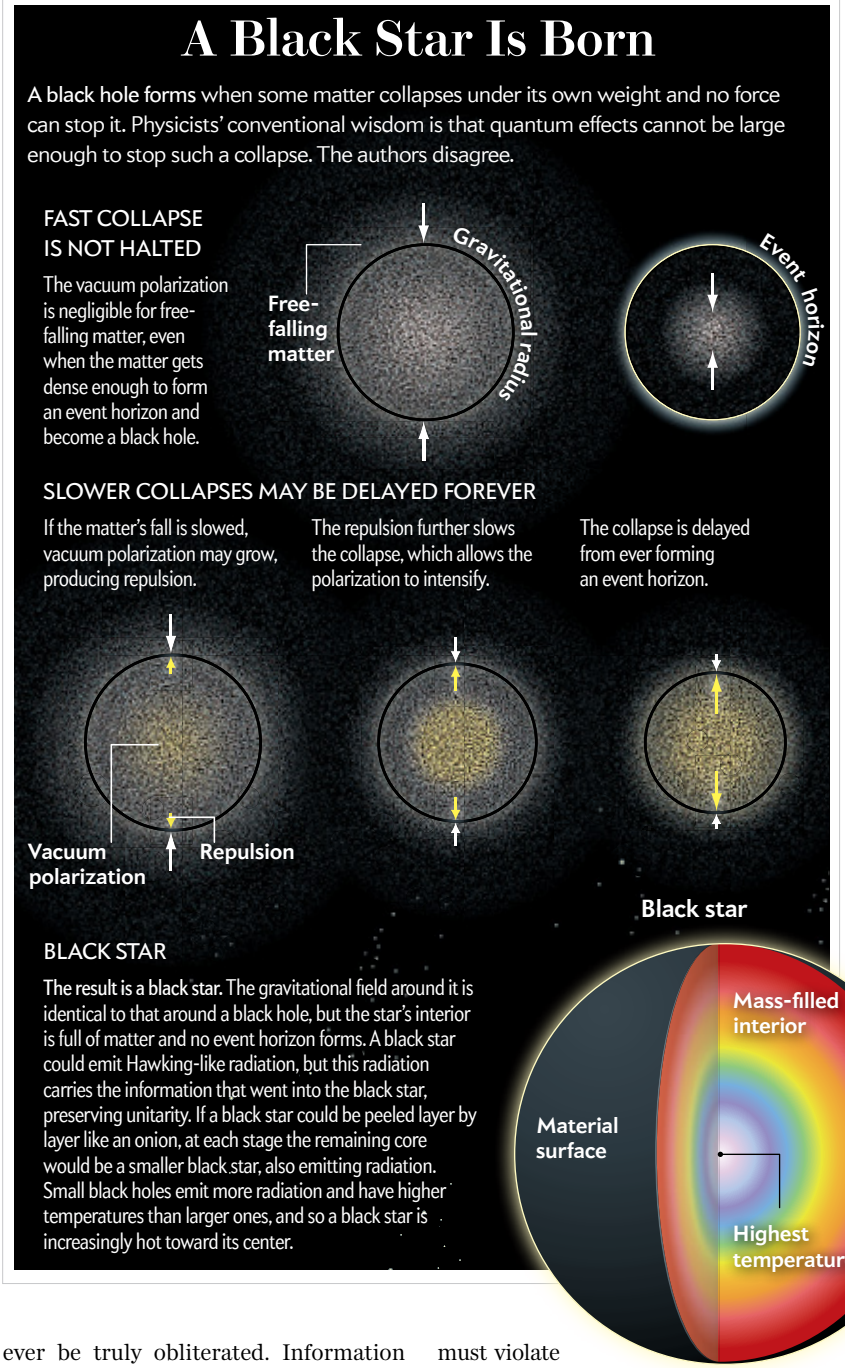
HAWKING'S DISCOVERY of black hole evaporation, along with earlier results by Jacob D. Bekenstein of the Hebrew University of Jerusalem, uncovered a deep—and as yet not fully understood—relation among gravity, quantum physics and thermodynamics. At the same time, it opened up new problems. Perhaps the most important is known as the information problem, which is closely related to the question of the final outcome of black hole evaporation.

Take the example of a large star undergoing gravitational collapse. The star embodies a vast amount of information in the positions and velocities and other properties of its more than 10^{55} particles. Suppose the star forms a black hole but then, gradually over the eons, evaporates by emitting Hawking radiation. A black hole's temperature is inversely proportional to its mass, and thus an evaporating black hole becomes hotter and evaporates faster as its mass and radius shrink. A huge explosion ejects the last of the black hole's mass. But what remains afterward? Does the hole completely vanish, or does some kind of small remnant remain? In either case, what has happened to all the information of the star? According to Hawking's calculation, the particles radiated by the hole carry essentially no information about the star's initial state. Even if some kind of black hole remnant remains, how could such a small object contain all the information that was in the original star?

The disappearance of information matters because one of the most fundamental pillars of quantum theory is that quantum states evolve in a manner that is called unitary, one consequence of which is that no information ought to ever be truly obliterated. Information may be inaccessible in practice, such as when an encyclopedia burns up, but in principle the information remains in the swirling smoke and ashes.

Because the calculations that predict Hawking radiation rely on semiclassical gravity, physicists cannot be sure if information loss is an artifact of the approximations involved or a feature that will remain when we discover how to compute the process exactly. If the evaporation process does destroy information, the correct full quantum gravity equations must violate the unitary nature of quantum mechanics as we know it. Conversely, if information is preserved and a complete theory of quantum gravity will reveal where it is in the radiation, either general relativity or quantum mechanics seems to need modification.

THE AUTHORS' PROPOSAL


A Black Star Is Born
A black hole forms when some matter collapses under its own weight and no force can stop it. Physicists' conventional wisdom is that quantum effects cannot be large enough to stop such a collapse. The authors disagree.
FAST COLLAPSE IS NOT HALTED
The vacuum polarization is negligible for free-falling matter, even when the matter gets dense enough to form an event horizon and become a black hole.
Free-falling matter
Gravitational radius
Event horizon
SLOWER COLLAPSES MAY BE DELAYED FOREVER
If the matter's fall is slowed, vacuum polarization may grow, producing repulsion.
The repulsion further slows the collapse, which allows the polarization to intensify.
The collapse is delayed from ever forming an event horizon.
Vacuum polarization
Repulsion
Black star
BLACK STAR
The result is a black star. The gravitational field around it is identical to that around a black hole, but the star's interior is full of matter and no event horizon forms. A black star could emit Hawking-like radiation, but this radiation carries the information that went into the black star, preserving unitarity. If a black star could be peeled layer by layer like an onion, at each stage the remaining core would be a smaller black star, also emitting radiation. Small black holes emit more radiation and have higher temperatures than larger ones, and so a black star is increasingly hot toward its center.
Mass-filled interior
Material surface
Highest temperature

A RADICAL ALTERNATIVE

THE INFORMATION PROBLEM and related puzzles have motivated us (and others) to revisit the line of reasoning that led

© 2013 Scientific American

physicists in the 1970s to the picture of evaporating almost classical black holes. We have found that the old semiclassical prediction that black holes form from gravitational collapse even when quantum effects are considered depends on several technical and often unstated assumptions.

In particular, the old calculations assume that collapse proceeds very rapidly, taking about the same time as would be needed for material at the star's surface to free-fall to the star's center. We found that for a slower collapse, quantum effects may produce a new kind of very compact object that does not have an event horizon and is thus much less problematic.

As we have already mentioned, the RSET of the quantum vacuum in a spacetime curved by a typical star is negligible everywhere. When the star starts to collapse, the RSET might change. Nevertheless, the old conclusion that the RSET remains negligible continues to hold if the collapse is about as fast as free falling.

Yet if the collapse proceeds significantly slower than free falling, the RSET can acquire arbitrarily large and negative values in the region near the Schwarzschild radius—where the classical event horizon would have formed. A negative RSET produces a repulsion, which further slows the collapse. The collapse might come to a complete halt just short of forming a horizon, or it might continue forever at an ever slower pace, becoming ever closer to forming a horizon but never actually producing one.

This result by itself, however, does not make it impossible for black holes to form. A perfectly homogeneous spherical cloud of matter of, say, 100 million solar masses falling freely under its own weight would surely produce a horizon. Such a large cloud would have a density comparable to that of water when it became compact enough to form a horizon. At such a low density we know no cause able to slow down the collapse. Accordingly, the RSET cannot become large enough to prevent a horizon from forming. But we know that what happened in the universe did not follow this script. The vast, nearly homogeneous clouds of matter that emerged from the early stages of the big bang did not collapse to form black holes. Instead a sequence of structures developed.

First, stars formed, the heat of their nuclear reactions delaying the collapse for a long time. When a star largely exhausts its nuclear fuel, it may develop into a white dwarf or, if massive enough, explode as a supernova, leaving behind a neutron star (a sphere made up of neutrons that is only somewhat larger than the star's gravitational radius). In either case, it is actually a purely quantum effect—the Pauli exclusion principle—that prevents further collapse. The neutrons in the

ALTERNATIVE BODIES

Other Ways out of a Hole

Many researchers have proposed more or less exotic objects that could serve as alternatives to the conventional (but apparently paradoxical) idea of an evaporating black hole and account for the dark, compact bodies observed by astronomers. The common feature of these proposals (and our own black star hypothesis) is that the new object would lack an event horizon.

GRAVASTARS

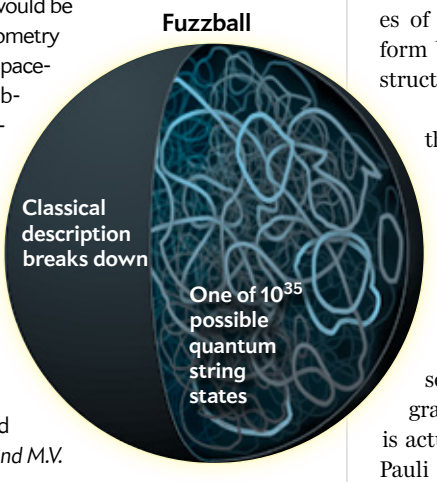
The spacetime geometry around a "*gra*vitational *va*cuum *star*" would be indistinguishable from that of a black hole down to about 10^{-35} meter away from the spherical region where the classical black hole horizon would have been located. The horizon would be replaced by a shell of matter and energy a mere 10^{-35} meter thick (known as the Planck length—the length scale at which quantum gravity effects are expected to become large). The gravastar's interior would be empty space with a large vacuum polarization, which would produce a repulsion that prevents the matter shell from collapsing any further. In a variant of the gravastar proposal, the classical notions of geometry break down in the region separating the interior and exterior.

BLACK HOLE COMPLEMENTARITY

In conventional quantum mechanics, complementarity refers to the idea that an observation may reveal either the particle nature of an object or the wave nature, but not both. Similarly, the quantum mechanics of black holes might embody a new kind of complementarity. An observer who remains outside a black hole may have a different description of its geometry than an observer who falls into the hole. This proposal is currently being revised; there are indications that complementarity could produce a firewall at the horizon, which would have to be consistently accommodated in the description of the resulting non-black hole object.

FUZZBALLS

Proponents of "fuzzballs" contend that the horizon would be a transition region between the exterior classical geometry and a quantum interior where no definite notion of spacetime could be specified. The interior would be describable by string theory and would not have a singularity (*right*). Each exterior geometry (say, the geometry of a black hole of exactly 10^{30} kilograms) could have any one of an exponentially large number of such stringy quantum states as its interior. The semiclassical view of a black hole—with an event horizon, an enormous entropy, a temperature and emission of thermal Hawking radiation—would amount to a statistical average over all the possible interiors, analogous to a description of a volume of gas that disregards the exact positions and motions of the individual atoms. *—C.B., S.L., S.S. and M.V.*



EXTREME PHYSICS

© 2013 Scientific American

neutron star cannot enter the same quantum state, and the resulting pressure resists the gravitational collapse. A similar story for ions and electrons explains why a white dwarf is stable.

If the neutron star acquires more mass, eventually the crushing gravitational load overwhelms the neutrons, and further collapse occurs. We do not know for certain what happens next (although the conventional view says a black hole forms). Scientists have suggested a variety of objects that might form—such as so-called quark stars, strange stars, boson stars and Q-balls—that would be stable at pressures too great for a neutron star. Physicists must develop a better understanding of how matter behaves at densities well beyond that of neutrons to know which conjecture, if any of them, is correct.

Thus, experience tells us that matter following the laws of quantum mechanics always seems to find new ways of delaying gravitational collapse. Although any of these roadblocks may be overcome (a typical stable configuration can always be made unstable by adding enough matter), each process that delays collapse provides additional time for the quantum vacuum's negative RSET to pile up and become significant. This RSET could take over the task of counterbalancing the gravitational pull, and because its repulsion may increase without limit, it can stop the matter's collapse to a black hole forever.

BLACK STARS

THE FINAL RESULTING BODIES would be the new kind of object we have named black stars. Because of their extremely small size and high density, they would share many observable properties with black holes, but conceptually they would be radically different. They would be material bodies, with a material surface and an interior filled with dense matter. They would be extremely dim because light emitted from their surface would be very redshifted—the light wave greatly stretched—in traveling from the intensely curved space near the black star to distant astronomers. In principle, astronomers could conduct complete astrophysical studies of black stars because no event horizon would present an obstacle.

Within the family of bodies of black star types, some might resemble evaporating black holes by emitting radiation similar to Hawking radiation. For the specific case in which collapse approaches formation of a horizon but never quite stops, we have shown that a black star could emit particles with a so-called Planckian energy spectrum (which is very similar to a thermal spectrum), at a temperature very slightly smaller than the Hawking temperature. By having no horizon, the black star cannot lock away any information. Instead the emitted particles and whatever matter remains behind with the black star carry all the information. Standard quantum physics would describe the formation and evaporation process. Black stars do not completely solve the information problem, however, as long as ways remain for event horizons to form somewhere in the universe.

These evaporating objects could be called quasi black holes because when viewed from the outside they would have approximately the same thermodynamic properties as evaporating black holes. Their interiors, however, would harbor a rainbow of temperatures, rising to a maximum near the center. If you imagine the body as an onionlike structure of concentric shells, each shell would be slowly shrinking, never quite compact enough for the combined mass of the shell and everything inside it to form a horizon. Each shell would be prevented from collapsing by the vacuum RSET that we predict will develop where the conditions for a horizon are approached slowly enough. The deeper shells would have higher temperatures, just like smaller-mass black holes do. We do not yet know whether these appealing objects show up naturally or whether they are exceptional.

What's Next

Future work on the black star scenario must show specific physical systems for which vacuum polarization succeeds in halting a collapse according to semiclassical gravity.

By describing quantum black holes as bundles of fundamental entities called branes, string theorists have reproduced predictions of semiclassical gravity for certain special cases. They hope to extend these results to all kinds of black holes.

A definitive resolution of the information problem and of the fate of collapsing matter will most likely require development of a complete quantum theory of gravity.

OVER THE HORIZON

STUDY OF BLACK HOLES has always provoked a great variety of reactions from researchers. On the one hand, it is exciting to think that they hide within them the door to unforeseeable new possibilities in physics, albeit only for those who dare to enter. On the other hand, implications of black holes have long disturbed some physicists—the quest for alternatives to black holes, often motivated by distaste for one or another of their features, is as old as the idea of black holes themselves.

Our black star proposal and other researchers' black hole alternatives all have the common theme that the spacetime around them is essentially identical to that around a classical black hole, down to extremely close to where the horizon itself would have formed. Although the secret door leading to an understanding of how quantum physics merges with gravity remains out of our sight, it may not be shielded from us by the impenetrable fortress of an event horizon. SA

Carlos Barceló, Stefano Liberati, Sebastiano Sonego and **Matt Visser** have been collaborating in various combinations and permutations since the turn of the millennium. Barceló is titular researcher of theoretical physics at the Institute of Astrophysics of Andalusia in Spain. Liberati is professor of astroparticle physics at the International School for Advanced Studies in Trieste, Italy. Sonego is professor of mathematical physics at the University of Udine in Italy. Visser is professor of mathematics at Victoria University of Wellington in New Zealand.

MORE TO EXPLORE

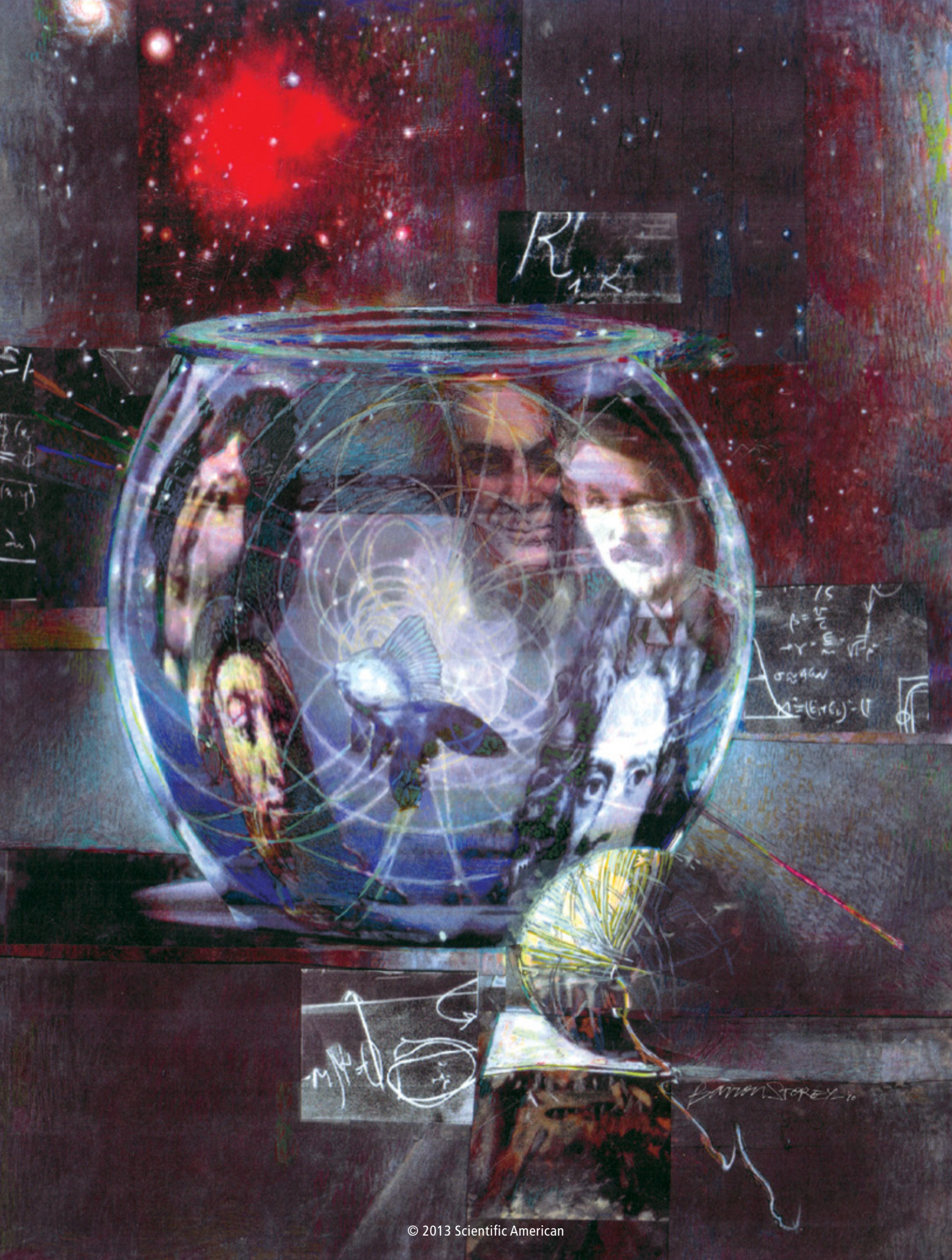
Fate of Gravitational Collapse in Semiclassical Gravity. Carlos Barceló, Stefano Liberati, Sebastiano Sonego and Matt Visser in *Physical Review D,* Vol. 77, No. 4; February 19, 2008.

Small, Dark, and Heavy: But Is It a Black Hole? Matt Visser, Carlos Barceló, Stefano Liberati and Sebastiano Sonego in *Proceedings of Black Holes in General Relativity and String Theory;* August 2008. Available at **http://arxiv.org/abs/0902.0346**

The Fuzzball Proposal for Black Holes. K. Skenderis and M. Taylor in *Physics Reports,* Vol. 467, Nos. 4–5, pages 117–171; October 2008. **http://arxiv.org/abs/0804.0552**

The Black Hole War: My Battle with Stephen Hawking to Make the World Safe for Quantum Mechanics. Leonard Susskind. Little Brown, 2008.

© 2013 Scientific American



© 2013 Scientific American

UNIFICATION

THE (ELUSIVE) THEORY OF EVERYTHING

Physicists have long sought to find one final theory that would unify all of physics. Instead they may have to settle for several

By Stephen Hawking and Leonard Mlodinow

ALMOST A DECADE AGO THE CITY COUNCIL OF MONZA, ITALY, barred pet owners from keeping goldfish in curved fishbowls. The sponsors of the measure explained that it is cruel to keep a fish in a bowl because the curved sides give the fish a distorted view of reality. Aside from the measure's significance to the poor goldfish, the story raises an interesting philosophical question: How do we know that the reality we perceive is true? The goldfish is seeing a version of reality that is different from ours, but can we be sure it is any less real? For all we know, we, too, may spend our entire lives staring out at the world through a distorting lens.

In physics, the question is not academic. Indeed, physicists and cosmologists are finding themselves in a similar predicament to the goldfish's. For decades we have strived to come up with an ultimate theory of everything—one complete and consistent set of fundamental laws of nature that explain every aspect of reality. It now appears that this quest may yield not a single theory but a family of interconnected theories, each describing its own version of reality, as if it viewed the universe through its own fishbowl.

This notion may be difficult for many people, including some working scientists, to accept. Most people believe that there is an objective reality out there and that our senses and our science directly convey information about the material world. Classical science is based on the belief that an external world exists whose properties are definite and independent of the observer who perceives them. In philosophy, that belief is called realism.

Those who remember Timothy Leary and the 1960s, however, know of another possibility: one's concept of reality can depend on the mind of the perceiver. That viewpoint, with various subtle differences, goes by names such as antirealism, instrumentalism or idealism. According to those doctrines, the world we know is constructed by the human mind employing sensory data as its raw material and is shaped by the interpretive structure of our brain. This viewpoint may be hard to accept, but it is not difficult to understand. There is no way to remove the observer—us—from our perception of the world.

Illustrations by Barron Storey

© 2013 Scientific American

The way physics has been going, realism is becoming difficult to defend. In classical physics—the physics of Newton that so accurately describes our everyday experience—the interpretation of terms such as object and position is for the most part in harmony with our commonsense, "realistic" understanding of those concepts. As measuring devices, however, we are crude instruments. Physicists have found that everyday objects and the light we see them by are made from objects—such as electrons and photons—that we do not perceive directly. These objects are governed not by classical physics but by the laws of quantum theory.

The reality of quantum theory is a radical departure from that of classical physics. In the framework of quantum theory, particles have neither definite positions nor definite velocities unless and until an observer measures those quantities. In some cases, individual objects do not even have an independent existence but rather exist only as part of an ensemble of many. Quantum physics also has important implications for our concept of the past. In classical physics, the past is assumed to exist as a definite series of events, but according to quantum physics, the past, like the future, is indefinite and exists only as a spectrum of possibilities. Even the universe as a whole has no single past or history. So quantum physics implies a different reality than that of classical physics—even though the latter is consistent with our intuition and still serves us well when we design things such as buildings and bridges.

These examples bring us to a conclusion that provides an important framework with which to interpret modern science. In our view, there is no picture- or theory-independent concept of reality. Instead we adopt a view that we call model-dependent realism: the idea that a physical theory or world picture is a model (generally of a mathematical nature) and a set of rules that connect the elements of the model to observations. According to model-dependent realism, it is pointless to ask whether a model is real, only whether it agrees with observation. If two models agree with observation, neither one can be considered more real than the other. A person can use whichever model is more convenient in the situation under consideration.

FRAMES OF REFERENCE

THE IDEA OF ALTERNATIVE REALITIES is a mainstay of today's popular culture. For example, in the science-fiction film *The Matrix* the human race is unknowingly living in a simulated virtual reality created by intelligent computers to keep them pacified and content while the computers suck their bioelectrical energy (whatever that is). How do we know we are not just computer-generated characters living in a Matrix-like world? If we lived in a synthetic, imaginary world, events would not necessarily have any logic or consistency or obey any laws. The aliens in control might find it more interesting or amusing to see our reactions, for example, if everyone in the world suddenly decided that chocolate was repulsive or that war was not an option, but that has never happened. If the aliens did enforce consistent laws, we would have no way to tell that another reality stood behind the simulated one. It is easy to call the world the aliens live in the "real" one and the computer-generated world a false one. But if—like us—the beings in the simulated world could not gaze into their universe from the outside, they would have no reason to doubt their own pictures of reality.

The goldfish are in a similar situation. Their view is not the same as ours from outside their curved bowl, but they could still formulate scientific laws governing the motion of the objects they observe on the outside. For instance, because light bends as it travels from air to water, a freely moving object that we would observe to move in a straight line would be observed by the goldfish to move along a curved path. The goldfish could formulate scientific laws from their distorted frame of reference that would always hold true and that would enable them to make predictions about the future motion of objects outside the bowl. Their laws would be more complicated than the laws in our frame, but simplicity is a matter of taste. If the goldfish formulated such a theory, we would have to admit the goldfish's view as a valid picture of reality.

A famous real-world example of different pictures of reality is the contrast between Ptolemy's Earth-centered model of the cosmos and Copernicus's sun-centered model. Although it is not uncommon for people to say that Copernicus proved Ptolemy wrong, that is not true. As in the case of our view versus that of the goldfish, one can use either picture as a model of the universe because we can explain our observations of the heavens by assuming either Earth or the sun to be at rest. Despite its role in philosophical debates over the nature of our universe, the real advantage of the Copernican system is that the equations of motion are much simpler in the frame of reference in which the sun is at rest.

Model-dependent realism applies not only to scientific models but also to the conscious and subconscious mental models we all create to interpret and understand the everyday world. For example, the human brain processes crude data from the optic nerve, combining input from both eyes, enhancing the resolution and filling in gaps such as the one in the retina's blind spot. Moreover, it creates the impression of three-dimensional space from the retina's two-dimensional data. When you see a chair, you have merely used the light scattered by the chair to build a mental image or model of the chair. The brain is so good at model building that if people are fitted with glasses that turn the images in their eyes upside down, their brain changes the model so

IN BRIEF

Stephen Hawking's work on black holes and the origin of the universe is arguably the most concrete progress theoretical physicists have made toward reconciling Einstein's gravitation and quantum physics into one final theory of everything.
Physicists have a favorite candidate for such a theory, string theory, but it comes in five different formulations, each covering a restricted range of situations.
A network of mathematical connections, however, links the different string theories into one overarching system, enigmatically called M-theory.
In their book *The Grand Design,* Hawking and California Institute of Technology physicist Leonard Mlodinow argue that the quest to discover a final theory may in fact never lead to a unique set of equations. Every scientific theory, they write, comes with its own model of reality, and it may not make sense to talk of what reality actually is. This essay is based on that book.

© 2013 Scientific American

that they again see things the right way up—hopefully before they try to sit down.

GLIMPSES OF THE DEEP THEORY

IN THE QUEST TO DISCOVER the ultimate laws of physics, no approach has raised higher hopes—or more controversy—than string theory. String theory was first proposed in the 1970s as an attempt to unify all the forces of nature into one coherent framework and, in particular, to bring the force of gravity into the domain of quantum physics. By the early 1990s, however, physicists discovered that string theory suffers from an awkward issue: there are five different string theories. For those advocating that string theory was the unique theory of everything, this was quite an embarrassment. In the mid-1990s researchers started discovering that these different theories—and yet another theory called supergravity—actually describe the same phenomena, giving them some hope that they would amount eventually to a unified theory. The theories are indeed related by what physicists call dualities, which are a kind of mathematical dictionaries for translating concepts back and forth. But, alas, each theory is a good description of phenomena only under a certain range of conditions—for example, at low energies. None can describe every aspect of the universe.

String theorists are now convinced that the five different string theories are just different approximations to a more fundamental theory called M-theory. (No one seems to know what the "M" stands for. It may be "master," "miracle" or "mystery," or all three.) People are still trying to decipher the nature of M-theory, but it seems that the traditional expectation of a single theory of nature may be untenable and that to describe the universe we must employ different theories in different situations. Thus, M-theory is not a theory in the usual sense but a network of theories. It is a bit like a map. To faithfully represent the entire Earth on a flat surface, one has to use a collection of maps, each of which covers a limited region. The maps overlap one another, and where they do, they show the same landscape. Similarly, the different theories in the M-theory family may look very different, but they can all be regarded as versions of the same underlying theory, and they all predict the same phenomena where they overlap, but none works well in all situations.

Whenever we develop a model of the world and find it to be successful, we tend to attribute to the model the quality of reality. But M-theory, like the goldfish example, shows that the same physical situation can be modeled in different ways, each employing different fundamental elements and concepts. It might be that to describe the universe we have to employ different theories in different situations. Each theory may have its own version of reality, but according to model-dependent realism, that diversity is acceptable, and none of the versions can be said to be more real than any other. It is not the physicist's traditional expectation for a theory of nature, nor does it correspond to our everyday idea of reality. But it might be the way of the universe. SA

Stephen Hawking's work laid the foundation for the modern understanding of black holes and the origin of the universe, although, as he has pointed out, he is at least as famous for his appearances on *The Simpsons* and *Star Trek: The Next Generation*. From 1979 to 2009, he was Lucasian Professor of Mathematics at the University of Cambridge, the post once held by Isaac Newton. His books include the classic *A Brief History of Time*, which has sold more than 10 million copies.

Leonard Mlodinow is a theoretical physicist at the California Institute of Technology. He has authored several books—including *Euclid's Window: The Story of Geometry from Parallel Lines to Hyperspace* and *The Drunkard's Walk: How Randomness Rules Our Lives*—as well as scripts for *MacGyver* and *Star Trek: The Next Generation*.

MORE TO EXPLORE

The Theory Formerly Known as Strings. Michael J. Duff in *Scientific American*, Vol. 278, No. 2, pages 64–69; February 1998.

The Illusion of Gravity. Juan Maldacena in *Scientific American*, Vol. 293, No. 5, pages 56–63; November 2005.

The Grand Design. Stephen Hawking and Leonard Mlodinow. Bantam Books, 2010.

© 2013 Scientific American

PARADOX

A QUANTUM THREAT TO SPECIAL RELATIVITY

Entanglement, like many quantum effects, violates some of our deepest intuitions about the world. It may also undermine Einstein's special theory of relativity

By David Z Albert and Rivka Galchen

OUR INTUITION, GOING BACK FOREVER, IS THAT TO MOVE, SAY, a rock, one has to touch that rock, or touch a stick that touches the rock, or give an order that travels via vibrations through the air to the ear of a person with a stick that can then push the rock—or some such sequence. This intuition, more generally, is that things can only directly affect other things that are right next to them. If A affects B *without* being right next to it, then the effect in question must be *in*direct—the effect in question must be something that gets transmitted by means of a chain of events in which each event brings about the next one directly, in a manner that smoothly spans the distance from A to B. Every time we think we can come up with an exception to this intuition—say, flipping a switch that turns on city street lights (but then we realize that this happens through wires) or listening to a BBC radio broadcast (but then we realize that radio waves propagate through the air)—it turns out that we have not, in fact, thought of an exception. Not, that is, in our everyday experience of the world.

IN BRIEF

In the universe as we experience it, we can directly affect only objects we can touch; therefore, the world seems local.

Quantum mechanics embraces action at a distance with a property called entanglement, in which two particles behave synchronously with no intermediary.

This nonlocal effect is not merely counterintuitive: it presents a serious problem to Einstein's special theory of relativity, thus shaking the foundations of physics.

Illustration by Jean-Francois Podevin

© 2013 Scientific American

© 2013 Scientific American

We term this intuition "locality."

Quantum mechanics has upended many an intuition but none deeper than this one. And this particular upending carries with it a threat, as yet unresolved, to special relativity—a foundation stone of our 21st-century physics.

THE THING FROM OUTER SPACE

LET'S BACK UP A BIT. Prior to the advent of quantum mechanics, and indeed back to the very beginnings of scientific investigations of nature, scholars believed that a complete description of the physical world could in principle be had by describing, one by one, each of the world's smallest and most elementary physical constituents. The full story of the world could be expressed as the sum of the constituents' stories.

Quantum mechanics violates this belief.

Real, measurable, physical features of collections of particles can, in a perfectly concrete way, exceed or elude or have nothing to do with the sum of the features of the individual particles. For example, according to quantum mechanics one can arrange a pair of particles so that they are precisely two feet apart and yet neither particle on its own has a definite position. Furthermore, the standard approach to understanding quantum physics, the so-called Copenhagen interpretation—proclaimed by the great Danish physicist Niels Bohr early last century and handed down from professor to student for generations—insists that it is not that we do not know the facts about the individual particles' exact locations; it is that there simply *aren't* any such facts. To ask after the position of a single particle would be as meaningless as, say, asking after the marital status of the number five. The problem is not epistemological (about what we know) but ontological (about what is).

Physicists say that particles related in this fashion are quantum mechanically entangled with one another. The entangled property need not be location: Two particles might spin in opposite ways, yet with neither one definitely spinning clockwise. Or exactly one of the particles might be excited, but neither is definitely the excited one. Entanglement may connect particles irrespective of where they are, what they are and what forces they may exert on one another—in principle, they could perfectly well be an electron and a neutron on opposite sides of the galaxy. Thus, entanglement makes for a kind of intimacy amid matter previously undreamt of.

Entanglement lies behind the new and exceedingly promising fields of quantum computation and quantum cryptography, which could provide the ability to solve certain problems that are beyond the practical range of an ordinary computer and the ability to communicate with guaranteed security from eavesdropping.

But entanglement also appears to entail the deeply spooky and radically counterintuitive phenomenon called nonlocality—the possibility of physically affecting something without touching it or touching any series of entities reaching from here to there. It's as if a fist in Des Moines could break a nose in Dallas without affecting any other physical thing (not a molecule of air, not an electron in a wire, not a twinkle of light) anywhere in the heartland.

The greatest worry about nonlocality, aside from its overwhelming intrinsic strangeness, has been that it intimates a profound threat to special relativity as we know it. In the past few years this old worry—finally allowed inside the house of serious thinking about physics—has become the centerpiece of debates that may finally dismantle, distort, reimagine, solidify or seed decay into the very foundations of physics.

RADICAL REVISIONS OF REALITY

ALBERT EINSTEIN had any number of worries about quantum mechanics. The overquoted concern about its chanciness ("God does not play dice") was just one. But the only objection he formally articulated, the only one he bothered to write a paper on, concerned the oddity of quantum-mechanical entanglement. This objection lies at the heart of what is now known as the EPR argument, named after its three authors, Einstein and his colleagues Boris Podolsky and Nathan Rosen [*see box on opposite page*]. In their 1935 paper "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?", they answer their own question with a tightly reasoned "no."

Their argument made pivotal use of one particular instruction in the quantum-mechanical recipe, or mathematical algorithm, for predicting the outcomes of experiments. Suppose that we measure the position of a particle that is quantum mechanically entangled with a second particle so that neither individually has a precise position, as we mentioned above. Naturally, when we learn the outcome of the measurement, we change our description of the first particle because we now know where it was for a moment. But the algorithm also instructs us to alter our description of the second particle—and to alter it instantaneously, no matter how far away it may be or what may lie between the two particles.

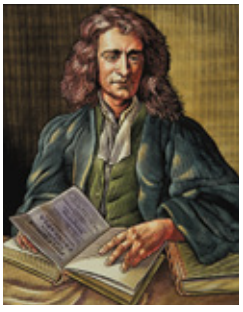
Entanglement was an uncontroversial fact of the picture of the world that quan-

B. SANDERSON *Science Source (Newton)*; THE GRANGER COLLECTION (*Coulomb apparatus*)

NONLOCALITY THROUGH THE CENTURIES

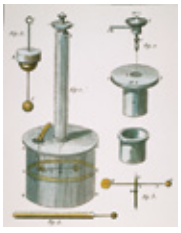
Changing Views of "Reality"

Our intuition is that the world is local: we can move a rock only by touching it directly, or by touching a stick that touches it, or by creating some unbroken chain of such direct, local connections. Yet since the beginnings of modern science in the 1600s, apparent nonlocalities have been challenging scientists.



1687: Isaac Newton's law of universal gravitation, the first modern scientific description of gravity, involves "action at a distance." Newton is sure there must be an account of gravity without this nonlocality and even tries an unsuccessful theory in which tiny invisible, jiggling particles fill all of seemingly empty space.

1785: Charles Coulomb introduces the inverse-square law for electrostatic forces, analogous to Newton's inverse-square law for gravity. Electric effects seem to involve action at a distance.



© 2013 Scientific American

tum mechanics presented to physicists, but it was a fact whose implications no one prior to Einstein had thought much about. He saw in entanglement something not merely strange but dubious. It struck him as spooky. It seemed, in particular, nonlocal.

Nobody at that time was ready to entertain the possibility that there were genuine physical nonlocalities in the world—not Einstein, not Bohr, not anybody. Einstein, Podolsky and Rosen took it for granted in their paper that the apparent nonlocality of quantum mechanics must be apparent only, that it must be some kind of mathematical anomaly or notational infelicity or, at any rate, that it must be a disposable artifact of the algorithm—surely one could cook up quantum mechanics's predictions for experiments without needing any nonlocal steps.

And in their paper they presented an argument to the effect that if (as everybody supposed) no genuine physical nonlocality exists in the world and if the experimental predictions of quantum mechanics are correct, then quantum mechanics must leave aspects of the world out of its account. There must be parts of the world's story that it fails to mention.

Bohr responded to the EPR paper practically overnight. His feverishly composed letter of refutation engaged none of the paper's concrete scientific arguments but instead took issue—in an opaque and sometimes downright oracular fashion—with its use of the word "reality" and its definition of "elements of physical reality." He talked at length about the distinction between subject and object, about the conditions under which it makes sense to ask questions and about the nature of human language. What science needed, according to Bohr, was a "radical revision of

BASICS

The EPR Thought Experiment

Albert Einstein, Boris Podolsky and Nathan Rosen ("EPR") pointed out that quantum entanglement of two particles produces strange results if two people who are far apart (*here Alice and Buzz*) each examine one of the particles.

Electrons have a property called spin, represented here by arrows that may point in any direction. When Alice measures an electron's spin (*below*), she chooses an axis. Measuring along a vertical axis, she will find the electron in either the up or the down state, with some probability for each. Using an east-west axis, she will find east or west spin.

Two particles may be entangled so that they have their spins pointing in opposite directions, even though neither has a definite direction of its own. Suppose that Alice and Buzz share such a pair and that Alice finds hers to have spin up (*below*). No matter how far away Buzz and his particle are from Alice, if he measures his particle along the vertical axis he will definitely see that his particle has spin down, the opposite of Alice's.

EPR argued that because Buzz can be 100 percent certain of measuring spin down, the spin of his particle must already be down, even prior to his measuring it. But Alice could equally well have measured along east-west and obtained, say, east spin, implying that Buzz's particle already has a west spin.

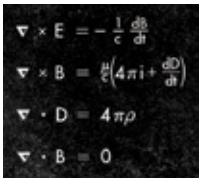
Because no quantum state permits Buzz's particle to be certainly spin down *and* certainly spin west, EPR concluded that quantum mechanics must be an incomplete theory.

CORDELIA MOLLOY *Science Source (Faraday's magnetic repulsion)*; LAWRENCE MANNING *Corbis (Fizeau and Foucault's speed of light)*; SCIENCE SOURCE *(Maxwell's equations)*

1831: Michael Faraday introduces the idea of magnetic lines of force. Physicists at this time use a notation involving electric and magnetic fields that fill space. The forces on a particle become, at least formally, a local action of the fields on them. But these fields are viewed as convenient calculational tools, not as things that are real.

1849: Hippolyte Fizeau and Jean-Bernard Foucault measure the speed of light to be 186,000 miles per second, or 298,000 kilometers per second, but no one knows what light really is.

1865: James Clerk Maxwell's equations reveal that electromagnetic fields have a rich, dynamical life of their own, pushing and pulling each other and crossing empty space at 298,000 km/s. Electromagnetism is local, and light is an electromagnetic wave!



Illustrations by Alfred T. Kamajian

© 2013 Scientific American

our attitude as regards physical reality."

Bohr did go out of his way to agree with the EPR paper on one point: that of course there can be no question of a genuine physical nonlocality. The apparent nonlocality, he argued, was just one more reason why we must abandon the quaint and outdated aspiration, so manifest in the EPR paper, of being able to read from the equations of quantum mechanics a realistic picture of the world—a picture of what actually exists before us from moment to moment. Bohr insisted, in effect, that not only do we see the world through a glass darkly but that this shadowy and indefinite view is as real as anything gets.

Bohr's was a curiously philosophical response to an explicitly scientific concern. More curious still was the enshrinement of Bohr's response as the official gospel of theoretical physics. To spend any more time on these matters became, thereafter, apostasy. The physics community thus turned away from its old aspirations to uncover what the world is really like, and for a long time thereafter it relegated metaphysical questions to the literature of fantasy.

Even today this crucial part of Einstein's legacy remains very much obscured. The best-selling 2007 Walter Isaacson biography of Einstein simply assures the reader that Einstein's criticism of quantum mechanics has since been resolved. And this is not true.

RETURN OF THE REPRESSED

THE FIRST SERIOUS scientific engagement with the EPR argument came (after 30 years of more or less complete neglect) in a famous 1964 paper by the extraordinary Irish physicist John S. Bell. From Bell's work it emerged that Bohr was wrong that nothing was wrong with his understanding of quantum mechanics and that Einstein was wrong about *what* was wrong with Bohr's understanding. To take in what was actually wrong involves abandoning the idea of locality.

The crucial question is whether the nonlocalities that at least appear to be present in the quantum-mechanical algorithm are merely apparent or something more. Bell seems to have been the first person to ask himself precisely what that question means. What could make genuine physical nonlocalities distinct from merely apparent ones? He reasoned that if any manifestly and completely local algorithm existed that made the same predictions for the outcomes of experiments as the quantum-mechanical algorithm does, then Einstein and Bohr would have been right to dismiss the nonlocalities in quantum mechanics as merely an artifact of that particular formalism. Conversely, if no algorithm could avoid nonlocalities, then they must be genuine physical phenomena. Bell then analyzed a specific entanglement scenario and concluded that no such local algorithm was mathematically possible.

And so the actual physical world is nonlocal. Period.

This conclusion turns everything upside down. Einstein, Bohr and everyone else had always taken it for granted that any genuine incompatibility between quantum mechanics and the principle of locality would be bad news for quantum mechanics. But Bell had now shown that locality was incompatible not merely with the abstract theoretical apparatus of quantum mechanics but with certain of its empirical predictions as well. Experimenters—in particular work by Alain Aspect of the Institute of Optics in Palaiseau, France, and his co-workers in 1981 and later—have left no doubt that those predictions are indeed correct. The bad news, then, was not for quantum mechanics but for the principle of locality—and thus, presumably, for special relativity, because it at least appears to rely on a presumption of locality [*see box on page 100*].

METAPHYSICAL MYSTERY TOUR

THE MAIN REACTION to Bell's work—one that persists in many quarters even today—was still more obfuscation. Bell had shown that any theory capable of reproducing the empirical predictions of quantum mechanics for entangled pairs of particles—including quantum mechanics itself—had to be genuinely physically nonlocal.

This message has been virtually ignored. Instead almost everyone says that what Bell showed is that any attempt at replacing the orthodox quantum-mechanical picture of the world with something more in tune with our classical metaphysical expectations—any so-called hidden-variable, deterministic or philosophically realist theory—would have to be nonlocal if it could reproduce the quantum-mechanical predictions for EPR systems [*for a couple of proposed escape clauses from Bell's conclusion, see box on page 101*]. People were at least reading Bell's work but as if through a convex looking glass.

Only a very small minority of physicists managed to avoid this particular misunderstanding and grasp that Bell's proof and Aspect's experiments meant the world itself had been discovered to be nonlocal, but even they almost universally believed that the nonlocality in question here posed no particular threat to special relativity.

This belief arises out of the idea that special relativity is inextricably bound up

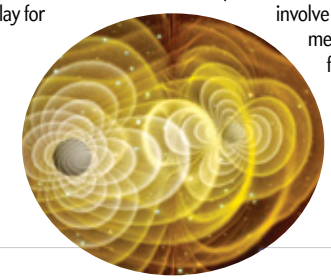
CORBIS (*Einstein*); COURTESY OF C. HENZE NASA (*curvature of spacetime*); PAUL EHRENFEST, COURTESY OF EMILIO SEGRÈ VISUAL ARCHIVES, EHRENFEST COLLECTION (*Einstein and Bohr*)

EXTREME PHYSICS

NONLOCALITY THROUGH THE CENTURIES

1905: Einstein's special theory of relativity reconciles Maxwell's equations with the principle that observers moving at a constant relative velocity should see identical laws of physics. But it destroys the possibility of distant events being simultaneous in any absolute sense.

1915: In Einstein's general theory of relativity, the curvature of spacetime plays the role that electromagnetic fields play for electromagnetic forces. Gravity is local: if a mass is jiggled, ripples in the curvature travel out at the speed of light.



1935: Einstein, Boris Podolsky and Nathan Rosen argue that because quantum-mechanical calculations involve nonlocal steps, quantum mechanics cannot be the full story. Niels Bohr (*far right*) insists we must accept quantum mechanics and instead give up old notions of "reality."

© 2013 Scientific American

with the impossibility of transmitting messages faster than the speed of light. After all, if special relativity is true, one can argue that no material carrier of a message can be accelerated from rest to speeds greater than that of light. And one can argue that a message transmitted faster than light would, according to some clocks, be a message that arrived before it was sent, potentially unleashing all the paradoxes of time travel.

As long ago as 1932 the brilliant Hungarian mathematician John von Neumann proved that the nonlocality of quantum mechanics cannot ever be parlayed into a mechanism to transmit messages instantaneously. For many decades virtually the entire theoretical physics community regarded von Neumann's proof as an assurance that quantum-mechanical nonlocality and special relativity can perfectly peacefully coexist.

VARIETIES OF NONLOCAL EXPERIENCE

IT TOOK YET ANOTHER 30 years after the publication of Bell's paper for physicists to finally look these issues squarely in the face. The first clear, sustained, logically flawless and uncompromisingly frank discussion of quantum nonlocality and relativity appeared in 1994, in a book with precisely that title by Tim Maudlin, now at New York University. His work highlighted how the compatibility of nonlocality and special relativity was a much more subtle question than the traditional platitudes based on instantaneous messages would have us believe.

Maudlin's work occurred against the backdrop of a new and profound shift in the intellectual environment. From the early 1980s onward, the grip of Bohr's conviction—that there could be no old-fashioned, philosophically realistic account of the subatomic world—was everywhere palpably beginning to weaken. By then a number of promising concrete scientific proposals seemed to provide a good account of that kind, at least in the approximation that neglects the effects of special relativity. These proposals included the Bohmian mechanics of David Bohm in England (developed in the early 1950s and an inspiration for Bell's work but otherwise largely ignored) and the GRW model of GianCarlo Ghirardi, Alberto Rimini and Tullio Weber in Italy. The old aspirations of physics to be a guide to metaphysics, to tell us literally and straightforwardly how the world actually is—aspirations that had lain dormant and neglected for more than 50 years—began, slowly, to reawaken.

THEORY MEETS REALITY

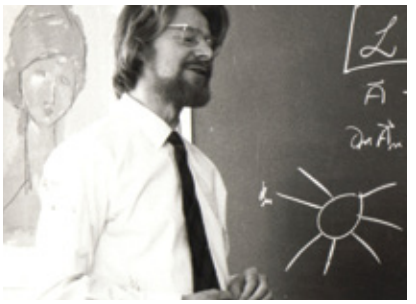
Bell's Theorem and the Physical World

The nonlocality of our physical world follows from a combination of a theorem proved by John S. Bell in 1964 and experimental results obtained since the early 1980s. His theorem builds on the puzzle about entangled particles pointed out by Einstein, Podolsky and Rosen in 1935 [*see box on page 97*]. The EPR argument assumes that nature is local so that in particular a measurement (by, say, Alice) on one particle of a widely separated entangled pair cannot instantaneously alter the physical state of the faraway partner particle (which, say, Buzz can measure). They conclude that Buzz's particle must already have determinate values for spins in every direction. Thus, quantum mechanics must be incomplete because it does not determine those values except to guarantee they will be consistent with whatever result Alice gets when she measures her particle.

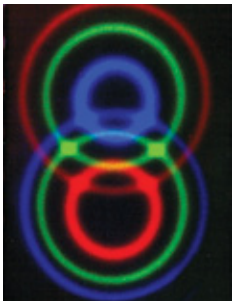
Bell asked: Supposing that Alice's and Buzz's entangled particles have determinate values, can such particles reproduce the results predicted by quantum mechanics for all the ways that Alice and Buzz might measure their particles? Recall that for particles with entangled spins, Alice and Buzz must each choose an axis to measure the spin along. Bell proved mathematically that if Alice and Buzz chose to measure along axes at angles such as 45 and 90 degrees from each other, their measurements from numerous runs of the experiment would produce a statistical distribution of results that disagreed with that predicted by quantum mechanics—no matter what distribution of determinate values the particles had.

Researchers carried out experiments using entangled photons instead of electrons (which alters the angles to use but makes the experiment technically much less difficult) and found results that conformed with quantum mechanics's predictions. And so by Bell's theorem there must not be any determinate values carried by those photons. And because that contradicts EPR's conclusion, the assumption that nature is local is also wrong. And so the universe we live in cannot be local. —*D.Z.A. and R.G.*

1964: John S. Bell (*right*) extends the "EPR" reasoning to cases in which spins are measured along nonparallel axes and shows that no local theory can possibly reproduce all of quantum mechanics's predictions for experimental results. The predictions of any local theory must always satisfy mathematical relations known as Bell's inequalities.



1981–present: Experiments using entangled states of light (*right*), in particular by Alain Aspect and his co-workers, verify that the world follows the predictions of quantum mechanics even in those situations in which quantum mechanics violates Bell's inequalities. The world is nonlocal after all.



CERN, COURTESY OF AIP EMILIO SEGRÈ VISUAL ARCHIVES (*Bell*); COURTESY OF PAUL KWIAT AND MICHAEL RECK *University of Vienna* (*entanglement*)

EXTREME PHYSICS

© 2013 Scientific American

Maudlin's book focused on three important points. First, the special theory of relativity is a claim about the geometric structure of space and time. The impossibility of transmitting mass or energy or information or causal influences faster than light—none of these requirements are even remotely, in and of themselves, sufficient to guarantee that the theory's claim about geometry is correct. Thus, von Neumann's proof about message transmission, in and of itself, offers us no assurance that quantum-mechanical nonlocality and special relativity can peacefully coexist.

Second, the truth of special relativity is (as a matter of fact) perfectly compatible with an enormous variety of hypothetical mechanisms for faster-than-light transmission of mass and energy and information and causal influence. In the 1960s, for example, Gerald Feinberg of Columbia University published an internally consistent and fully relativistic theory of a hypothetical species of particle—tachyons—for which it is physically impossible ever to travel *slower* than light. Maudlin invented other examples.

Thus, the mere existence of a nonlocality in quantum mechanics, in and of itself, does not mean that quantum mechanics cannot coexist with special relativity. So perhaps there is hope.

As Maudlin emphasized in his third point, however, the particular variety of action at a distance that we encounter in quantum mechanics is an entirely different animal from the kind exemplified by Feinberg's tachyons or Maudlin's other examples. What is uncanny about the way that quantum-mechanical particles can nonlocally influence one another is that it does not depend on the particles' spatial arrangements or their intrinsic physical characteristics—as all the relativistic influences alluded to in the preceding paragraphs do—but only on whether or not the particles in question are quantum mechanically entangled with one another.

The kind of nonlocality one encounters in quantum mechanics seems to call for an absolute simultaneity, which would pose a very real and ominous threat to special relativity.

That's the rub.

HOPE FOR SPECIAL RELATIVITY?

TWO NEW RESULTS—pulling in curiously different directions—have emerged from this discussion in just the past few years. The first suggests a way that quantum-mechanical nonlocality could be compatible with special relativity; the other reveals a new blow that the combination of quantum mechanics and special relativity strikes against our deepest intuitions of the world.

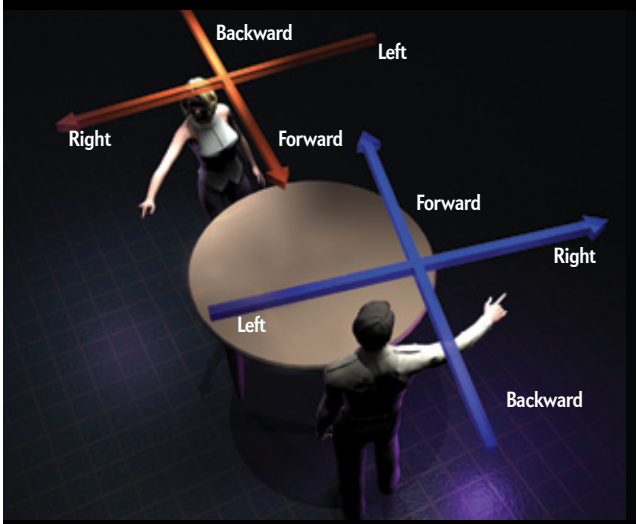
The first result appeared in an astonishing 2006 paper by Roderich Tumulka, a young German mathematician now at Rutgers University. Tumulka showed how all the empirical predictions of quantum

THE PROBLEM

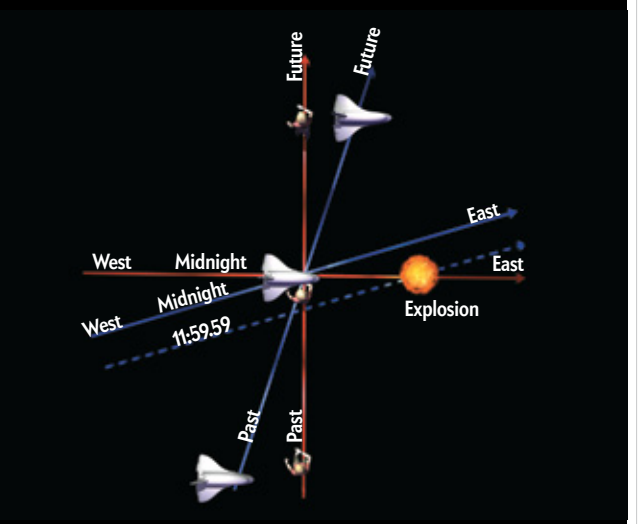
Why Special Relativity Dislikes Nonlocality

The special theory of relativity reveals an essential geometric relation between space and time that had never previously been imagined. The relation makes the concept of "instantaneous action at a distance" not merely strange but downright unintelligible.

Here Alice and Buzz cannot agree about which distant events are simultaneous, nor can they agree about a theory involving action at a distance, such as one in which Alice causes the distant explosion "instantaneously" by pressing a button at midnight.



Alice and Buzz, who are standing at different places around a table, disagree about the spatial directions "right," "left," "forward" and "backward." Special relativity shows that people in relative motion disagree about time as well as space.



Alice's space and time axes (*red*) are centered on where she stands at precisely midnight. Buzz flies over Alice, heading east at nearly the speed of light. His motion tilts his space and time axes (*blue*) relative to Alice's. The duo disagree about when a bomb exploded several kilometers away: Alice insists it happened at midnight, but Buzz says it happened a second earlier (*blue dashed line*).

© 2013 Scientific American

mechanics for entangled pairs of particles could be reproduced by a clever modification of the GRW theory (recall that this theory proposes a philosophically realist way to get the predictions of quantum mechanics under many circumstances). The modification is nonlocal, and yet it is fully compatible with the spacetime geometry of special relativity.

This work is still very much in its infancy. No one has yet been able to write down a satisfactory version of Tumulka's theory that can be applied to particles that attract or repel one another. Moreover, his theory introduces a new variety of nonlocality into the laws of nature—a nonlocality not merely in space but in time! To use his theory to determine the probabilities of what happens next, one must plug in not only the world's current complete physical state (as is customary in a physical theory) but also certain facts about the past. That feature and some others are worrying, but Tumulka has certainly taken away some of the grounds for Maudlin's fear that quantum-mechanical nonlocality cannot be made to peacefully coexist with special relativity.

The other recent result, discovered by one of us (Albert), showed that combining quantum mechanics and special relativity requires that we give up another of our primordial convictions. We believe that everything there is to say about the world can in principle be put into the form of a narrative, or story. Or, in more precise and technical terms: everything there is to say can be packed into an infinite set of propositions of the form "at *t*1 *this* is the exact physical condition of the world" and "at *t*2 *that* is the exact physical condition of the world," and so on. But the phenomenon of quantum-mechanical entanglement and the spacetime geometry of special relativity—taken together—imply that the physical history of the world is infinitely too rich for that.

That result, like most theoretical results in quantum mechanics, involves manipulating and analyzing a mathematical entity called a wave function, a concept Erwin Schrödinger introduced eight decades ago to define quantum states. It is from wave functions that physicists infer the possibility (indeed, the necessity) of entanglement, of particles having indefinite positions, and so forth. And it is the wave function that lies at the heart of puzzles about the nonlocal effects of quantum mechanics.

But what *is* it, exactly? Investigators of the foundations of physics are now vigorously debating that question. Is the wave function a concrete physical object, or is it more like a law of motion or an internal property of particles or a relation among spatial points? Or is it merely our current information about the particles? Or *what?*

Quantum-mechanical wave functions cannot be represented mathematically in anything smaller than a mind-bogglingly high-dimensional space called a configuration space. If, as some argue, wave functions need to be thought of as concrete physical objects, then we need to take seriously the idea that the world's history plays itself out not in the three-dimensional space of our everyday experience or the four-dimensional spacetime of special relativity but rather this gigantic and unfamiliar configuration space, out of which the illusion of three-dimensionality somehow emerges. Our three-dimensional idea of locality would need to be understood as emergent as well. The nonlocality of quantum physics might be our window into this deeper level of reality.

The status of special relativity, just more than a century after it was presented to the world, is suddenly a radically open and rapidly developing question. This situation has come about because physicists and philosophers have finally followed through on the loose ends of Einstein's long-neglected argument with quantum mechanics—an irony-laden further proof of Einstein's genius. He may turn out to have been wrong just where he was widely thought to be right and right just where he was widely thought to be wrong. SA

Other Ways Out

Some physicists argue that John S. Bell's mathematical proof of the nonlocality of the quantum-mechanical world has some escape clauses.

MANY WORLDS

Bell innocently assumes that quantum experiments have unique outcomes. The many-worlds interpretation, however, asserts that quantum measurements in effect split the universe into branches where all the different outcomes occur in parallel. So your universe can be "local" if copies of you inhabit myriad unseen parallel universes. This approach, however, is beset by many difficult problems.

REALISM?

Many believe that because Bell starts by assuming the world conforms to what is called local realism, he therefore proved that *either* locality or realism is violated. Thus, the world could be local if it violates "realism." But this idea overlooks—or misunderstands—that the original "EPR" argument of Albert Einstein, Boris Podolsky and Nathan Rosen rules out the possibility of quantum locality without the realism that Bell uses. *— D.Z.A. and R.G.*

David Z Albert and **Rivka Galchen** both teach at Columbia University, one on how physics tells the story of the world, the other on how to write stories. Albert is Frederick E. Woodbridge Professor of Philosophy at Columbia and author of *Quantum Mechanics and Experience* and *Time and Chance.* Galchen is an adjunct professor in the writing division of Columbia's School of the Arts. Her often science-steeped stories and essays have appeared in the *New Yorker,* the *New York Times* and the *Believer.* Her first novel, *Atmospheric Disturbances,* was published by Farrar, Straus and Giroux in 2008.

MORE TO EXPLORE

Quantum Theory and Measurement. Edited by John Archibald Wheeler and Wojciech Hubert Zurek. Princeton University Press, 1983. (Includes the original "EPR" paper and Niels Bohr's response.)

Quantum Mechanics and Experience. David Z Albert. Harvard University Press, 1992.

The Shaky Game: Einstein, Realism, and the Quantum Theory. Second edition. Arthur Fine. University of Chicago Press, 1996.

Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics. Second edition. Tim Maudlin. Wiley-Blackwell, 2002.

Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy. Second edition. J. S. Bell. Cambridge University Press, 2004.

© 2013 Scientific American

© 2013 Scientific American

LIMITS

BEYOND THE QUANTUM HORIZON

Once viewed as imposing absolute limits on knowledge and technology, quantum theory is now expanding the power of computers and the vistas of the mind

By David Deutsch and Artur Ekert

LATE IN THE 19TH CENTURY AN UNKNOWN ARTIST DEPICTED A TRAVELER who reaches the horizon, where the sky meets the ground. Kneeling in a stylized terrestrial landscape, he pokes his head through the firmament to experience the unknown [*see illustration on next page*]. The image, known as the Flammarion engraving, illustrates the human quest for knowledge. Two possible interpretations of the visual metaphor correspond to two sharply different conceptions of knowledge.

Either it depicts an *imaginary* barrier that, in reality, science can always pass through, or it shows a *real* barrier that we can penetrate only in our imagination. By the latter reading, the artist is saying that we are imprisoned inside a finite bubble of familiar objects and events. We may expect to understand the world of direct experience, but the infinity outside is inaccessible to exploration and to explanation. Does science continually transcend the familiar and reveal new horizons, or does it show us that our prison is inescapable—teaching us a lesson in bounded knowledge and unbounded humility?

Quantum theory is often given as the ultimate argument for the latter vision. Early on, its theorists developed a tradition of gravely teaching willful irrationality to students: "If you think you understand quantum theory, then you don't." "You're not allowed to ask that question." "The theory is inscrutable and so, therefore, is the world." "Things happen without reason or explanation." So textbooks and popular accounts have typically said.

Yet the developments of the past couple of decades contradict those characterizations. Throughout the history of the field, physicists often assumed that various kinds of constraints from quantum physics would prevent us from fully harnessing nature in the way that classical

Illustration by Goñi Montes

© 2013 Scientific American

mechanics had accustomed us to. None of these impediments have ever materialized. On the contrary, quantum mechanics has been liberating. Fundamentally, quantum-mechanical attributes of objects, such as superposition, entanglement, discreteness and randomness, have proved not to be limitations but resources. Using them, inventors have fashioned all kinds of miraculous devices, such as lasers and microchips.

These were just the beginning. We will increasingly use quantum phenomena for communications and computation systems that are unfathomably powerful from a classical point of view. We are discovering novel ways of harnessing nature and even of creating knowledge.

FLAMMARION ENGRAVING: **This famous 19th-century wood engraving (first printed in black and white) poses the question: Is knowledge bounded, or can we always poke our head into the beyond?**

BEYOND UNCERTAINTY

IN 1965 INTEL co-founder Gordon Moore predicted that engineers would double the number of transistors on a chip every two years or so. Now known as Moore's law, this prediction has held true for more than half a century. Yet from the outset, it rang warning bells. If the law continued to hold, you could predict when transistors would reach the size of individual atoms—and then what? Engineers would enter the realm of the unknowable.

In the traditional conception of quantum theory, the uncertainty principle sets a limit that no technological progress could ever overcome: the more we know about some properties, such as a particle's position, the less we can know about others, such as the particle's speed. What cannot be known cannot be controlled. Attempts to manipulate tiny objects meet with rampant randomness, classically impossible correlations, and other breakdowns of cause and effect. An inescapable conclusion followed: the end of progress in information technology was nigh.

Today, however, physicists routinely exert control over the quantum world without any such barrier. We encode information in individual atoms or elementary particles and process it with exquisite precision, despite the uncertainty principle, often creating functionality that is not achievable in any other way. But how?

Let us take a closer look at a basic chunk of information, as traditionally conceived: the bit. To a physicist, a bit is a physical system that can be prepared in one of two different states, representing two logical values: no or yes, false or true, 0 or 1. In digital computers, the presence or absence of a charge on the plates of a capacitor can represent a bit. At the atomic level, one can use two states of an electron in an atom, with 0 represented by the lowest-energy (ground) state and 1 by some higher-energy state.

To manipulate this information, physicists shine pulses of light on the atom. A pulse with the right frequency, duration and amplitude, known as a π-pulse, takes state 0 into state 1, and vice versa. Physicists can adjust the frequency to manipulate two interacting atoms, so that one atom controls what happens to the other. Thus, we have all the ingredients for one- and two-bit logic gates, the building blocks of classical computers, without any impediment from the uncertainty principle.

To understand what makes this feat of miniaturization possible, we have to be clear about what the uncertainty principle does and does not say. At any instant, some of the properties of an atom or other system, called its observables, may be "sharp"—possess only one value at that instant. The uncertainty principle does not rule out sharp observables. It merely states that not all observables in a physical system can be sharp at the same time. In the atom example, the sharp observable is energy: in both the 0 and 1 states, the electron has a perfectly well-defined energy. Other observables, such as position and velocity, are not sharp; the electron is delocalized, and its velocity likewise takes a range of different values simultaneously. If we attempted to store information using position and velocity, we would indeed encounter a quantum limit. The answer is not to throw up our hands in despair but to make a judicious choice of observables to serve as computer bits.

This situation recalls the comedy routine in which a patient tells a doctor, "It hurts when I do this," to which the doctor replies, "Don't do that." If some particle properties are hard to make sharp, there is a simple way around that: do not attempt to store information in those properties. Use some other properties instead.

BEYOND BITS

IF ALL WE WANT TO DO is build a classical computer using atoms rather than transistors as building blocks, then sharp observables are all we need. But quantum mechanics offers much more. It allows us to make powerful use of nonsharp observables, too. The fact that observables can take on multiple values at the same time greatly enriches the possibilities.

For instance, energy is usually a sharp observable, but we can turn it into a nonsharp one. In addition to being in its ground state or its excited state, an electron in an atom can also be in a superposition—both states at once. The electron is still in a perfectly definite state, but instead of being either 0 or 1, it is 0 *and* 1.

IN BRIEF

Quantum mechanics used to be described as a theory of limits, implying that our observations are unavoidably uncertain, that randomness rules the world, and that the theory itself is too weird to master and forces us to abandon the very idea that there is a world out there that science could accurately explain and describe.

Those misconceptions are rooted in philosophical doctrines, such as logical positivism, that were popular during the period in the early 20th century when physicists developed and honed the theory.

In truth, quantum mechanics imposes no significant limits on knowledge. The quantum world has a richness and intricacy that allow for new practical technologies, including powerful computers.

CORBIS

© 2013 Scientific American

Any physical object can do this, but an object in which such states can be reliably prepared, measured and manipulated is called a quantum bit, or qubit. Pulses of light can make the energy of an electron change not only from one sharp value to another but from sharp to nonsharp, and vice versa. Whereas a π-pulse swaps states 0 and 1, a pulse of the same frequency but half the duration or amplitude, known as a $\pi/2$-pulse, sends the electron to a superposition of 0 and 1.

If we attempted to measure the energy of the electron in such a superposition, we would find it was either the energy of the ground state or the energy of the excited state with equal probability. In that case, we would encounter randomness, just as the naysayers assert. Once again, we can readily sidestep this apparent roadblock—and in doing so create radically new functionality. Instead of measuring the electron in this superposition, we leave it there. For instance, start with an electron in state 0, send in a $\pi/2$-pulse, then send in a second $\pi/2$-pulse. Now measure the electron. It will be in state 1 with a 100 percent probability [*see box on next page*]. The observable is sharp once again.

To see the significance, consider the most basic logic gate in a computer, NOT. Its output is the negation of the input: 0 goes to 1, 1 to 0. Suppose you were given the following assignment: design the square root of NOT—that is, a logic gate that, acting twice in succession on an input, negates it. Using only classical equipment, you would find the assignment impossible. Yet a $\pi/2$-pulse implements this "impossible" logic gate. Two such pulses in succession have exactly the desired effect. Two consecutive applications of the square root of NOT convert 0 to 1, or 1 to 0.

Note that the intermediate superpositions of 0 and 1 are not the same in the two cases. They both contain the same proportion of 0 and 1 but differ in another variable (called their relative phase). Experimental physicists have built this logic gate and other classically impossible gates using qubits made of such things as photons, trapped ions, atoms and nuclear spins. They are the building blocks of a quantum computer.

BEYOND CLASSICAL COMPUTATION

TO SOLVE a particular problem, computers (classical or quantum) follow a pre-

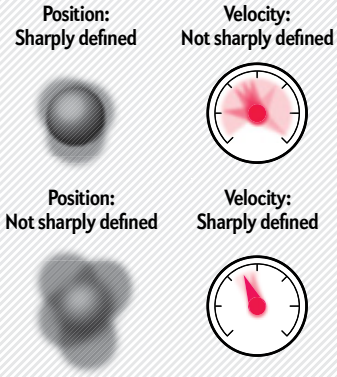
SHOWSTOPPERS NO LONGER

Supposed Limits to Quantum Computing—and How to Break Them

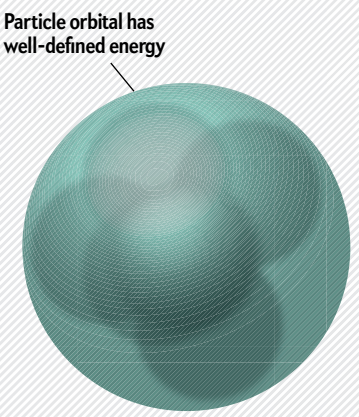
Quantum mechanics is often portrayed as the ultimate obstacle to the miniaturization of electronics. Fortunately, it is no such thing. Physicists have learned to work around the barriers they used to worry about. In fact, it is at the quantum level that computers will reach their true potential, achieving a power far beyond that of ordinary machines.

Uncertainty Principle

PROBLEM: The famous Heisenberg uncertainty principle limits the precision of certain measurements. If you pin down the position of a particle exactly, it will start moving with a range of different velocities simultaneously; if you measure its velocity exactly, you likewise force its position to spread out uncontrollably. Therefore, these properties are unreliable ways to store information.

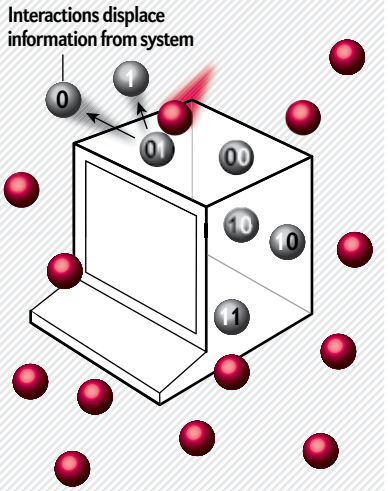


SOLUTION: Not all quantum measurements are subject to this limitation. In situations where position and velocity are uncertain, other properties such as energy may be perfectly well defined. In situations where energy is uncertain, some other variables may be suitable.

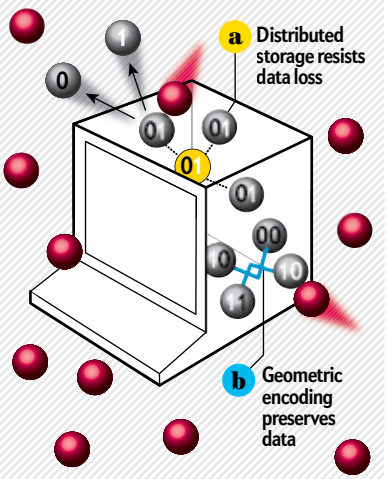


Decoherence

PROBLEM: The particles that make up a computer interact with the surroundings, so that information spreads out, spoiling quantum computations.



SOLUTION: Error-correction procedures can compensate for decoherence long enough to complete a computation. For instance, physicists can spread quantum information over multiple particles **a** or encode it in a geometric form that is naturally resistant to noise **b**.



Illustrations by Jen Christiansen

© 2013 Scientific American

cise set of instructions—an algorithm. Computer scientists quantify the efficiency of an algorithm according to how rapidly its running time increases when it is given ever larger inputs to work on. For example, using the algorithm taught in elementary school, one can multiply two n-digit numbers in a time that grows like the number of digits squared, n^2. In contrast, the fastest-known method for the reverse operation—factoring an n-digit integer into prime numbers—takes a time that grows exponentially, roughly as 2^n. That is considered inefficient.

By providing qualitatively new logic gates, quantum mechanics makes new algorithms possible. One of the most impressive examples is for factoring. A quantum algorithm discovered in 1994 by Peter Shor, then at Bell Laboratories, can factor n-digit numbers in a series of steps that grows only as n^2. For other problems, such as searching a long list, quantum computers offer less dramatic but nonetheless significant advantages. To be sure, not all quantum algorithms are so efficient; many are no faster than their classical counterparts.

Most likely, the first practical applications of general-purpose quantum computers will not be factorization but the simulation of other quantum systems—a task that takes an exponentially long time with classical computers. Quantum simulations may have a tremendous impact in fields such as the discovery of new drugs and the development of new materials.

Skeptics of the practicality of quantum computing cite the arduous problem of stringing together quantum logic gates. Apart from the technical difficulties of working at single-atom and single-photon scales, the main problem is that of preventing the surrounding environment from spoiling the computation. This process, called decoherence, is often presented as a fundamental limit to quantum computation. It is not. Quantum theory itself provides the means of correcting errors caused by decoherence. If the sources of error satisfy certain assumptions that can plausibly be met by ingenious designers—for instance, that the random errors occur independently on each of the qubits and that the logic gates are sufficiently accurate—then quantum computers can be made fault-tolerant. They can operate reliably for arbitrarily long durations.

BEYOND CONVENTIONAL MATHEMATICAL KNOWLEDGE

THE STORY of the "impossible" logic gates illustrates a startling fact about the physics of computation. When we improve our knowledge about physical reality, we sometimes improve our knowledge of the abstract realms of logic and mathematics, too. Quantum mechanics will transform these realms as surely as it already has transformed physics and engineering.

The reason is that although mathematical *truths* are independent of physics, we acquire *knowledge* of them through physical processes, and which ones we can know depends on what the laws of physics are. A mathematical proof is a sequence of logical operations. So what is provable and not provable depends on what logical operations (such as NOT) the laws of physics allow us to implement. These operations must be so simple, physically, that we know, without further proof, what it means to perform them, and that judgment is rooted in our knowledge of the physical world. By expanding our repertoire of such elementary computations to include ones such as the square root of NOT, quantum physics will allow mathematicians to poke their heads through a barrier previously as-

EXTREME PHYSICS

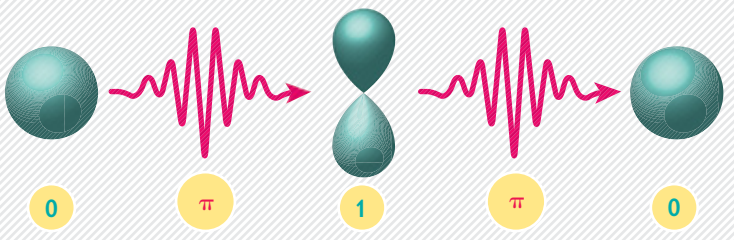
QUANTUM LOGIC

Impossible … NOT!

Quantum computers not only can do anything a classical computer can but also can perform operations outside the scope of classical logic. In this example, two energy states of an electron in an atom represent the 0 and 1 of a computer bit. In both states, the electron has no specific position and velocity: it is spread out over spherical and oval regions called orbitals, and its velocity takes a range of different values simultaneously. Nevertheless, the two states have different energies, and it is the energy that determines the bit value.

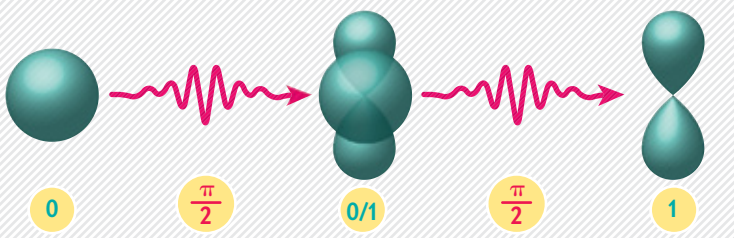
Ordinary NOT

To perform the most basic computational operation, NOT, which inverts the value of a bit, physicists shine pulses of light of appropriate frequency, duration and intensity—known as π-pulses—on the atom. If the electron begins in the 0 state, it will end up in the 1, and vice versa.



Square Root of NOT

The same procedure can be modified to perform a seemingly impossible computational operation: the square root of NOT. A so-called $\pi/2$-pulse, with a lesser amplitude or shorter duration than the π-pulse, sends the electron from the 0 or 1 state into a combination, or superposition, of both states. A second $\pi/2$-pulse then bumps the electron into either the 1 state (if it started as 0) or the 0 state (if it started as 1). This and other new operations give quantum computers their immense power.



© 2013 Scientific American

sumed to exist in the world of pure abstractions. They will be able to see, and to prove, truths there that would otherwise remain hidden forever.

For example, suppose the answer to some unsolved mathematical puzzle depends on knowing the factors of some particular enormous integer N—so enormous that even if all the matter in the universe were made into classical computers that then ran for the age of the universe, they would still not be able to factor it. A quantum computer could do so quickly. When mathematicians publish the solution, they will have to state the factors at the outset, as if pulled out of a magician's hat: "Here are two integers whose product is N." No amount of paper could ever suffice to detail how they had obtained those factors.

In this way, a quantum computer would supply the essential key that solves the mathematical puzzle. Without that key, which no classical process could realistically provide, the result would never be known. Some mathematicians already consider their subject an empirical science, obtaining its results not only by careful reasoning but also by experiments. Quantum physics takes that approach to a new level and makes it compulsory.

BEYOND BAD PHILOSOPHY

IF QUANTUM MECHANICS allows new kinds of computation, why did physicists ever worry that the theory would limit scientific progress? The answer goes back to the formative days of the theory.

Erwin Schrödinger, who discovered quantum theory's defining equation, once warned a lecture audience that what he was about to say might be considered insane. He went on to explain that when his famous equation describes different histories of a particle, those are "not alternatives but all really happen simultaneously." Eminent scientists going off the rails is not unknown, but this 1933 Nobelist was merely making what should have been a modest claim: that the equation for which he had been awarded the prize was a true description of the facts. Schrödinger felt the need to be defensive not because he had interpreted his equation irrationally but precisely because he had not.

How could such an apparently innocuous claim ever have been considered outlandish? It was because the majority of physicists had succumbed to bad philosophy: philosophical doctrines that actively hindered the acquisition of other knowledge. Philosophy and fundamental physics are so closely connected—despite numerous claims to the contrary from both fields—that when the philosophical mainstream took a steep nosedive during the first decades of the 20th century, it dragged parts of physics down with it.

The culprits were doctrines such as logical positivism ("If it's not verifiable by experiment, it's meaningless"), instrumentalism ("If the predictions work, why worry about what brings them about?") and philosophical relativism ("Statements can't be objectively true or false, only legitimized or delegitimized by a particular culture"). The damage was done by what they had in common: denial of realism, the commonsense philosophical position that the physical world exists and that the methods of science can glean knowledge about it.

Physics is dragging philosophy back on track. People want to understand reality, no matter how loudly they deny that.

It was in that philosophical atmosphere that physicist Niels Bohr developed an influential interpretation of quantum theory that denied the possibility of speaking of phenomena as existing objectively. One was not permitted to ask what values physical variables had while not being observed (such as halfway through a quantum computation). Physicists who, by the nature of their calling, could not help wanting to ask tried not to. Most of them went on to train their students not to. The most advanced theory in the most fundamental of the sciences was deemed to be stridently contradicting the very existence of truth, explanation and physical reality.

Not every philosopher abandoned realism. Bertrand Russell and Karl Popper were notable exceptions. Not every physicist did, either. Albert Einstein and David Bohm bucked the trend, and Hugh Everett proposed that physical quantities really do take on more than one value at once (the view we ourselves endorse). On the whole, however, philosophers were uninterested in reality, and although physicists went on using quantum theory to study other areas of physics, research on the nature of quantum processes themselves lost its way.

Things have been gradually improving for a couple of decades, and it has been physics that is dragging philosophy back on track. People want to understand reality, no matter how loudly they may deny that. We are finally sailing past the supposed limits on knowledge that bad philosophy once taught us to resign ourselves to.

What if the theory is eventually refuted—if some deeper limitation foils the attempt to build a scalable quantum computer? We would be thrilled to see that happen. Such an outcome is by far the most desired one. Not only would it lead to a revision of our fundamental knowledge about physics, we would expect it to provide even more fascinating types of computation. For if something stops quantum mechanics, we shall expect to have an exciting new whatever-stops-quantum-mechanics theory, followed by exciting new whatever-stops-quantum-computers computers. One way or another, there will be no limits on knowledge or progress. SA

David Deutsch, University of Oxford physicist and inventor of the concept of universal quantum computers, says he got interested in physics as a child when he rebelled at the claim that no one can understand everything that is understood.

Artur Ekert pioneered entanglement-based cryptography as a graduate student. He is now director of the Center for Quantum Technologies in Singapore and a professor at Oxford's Mathematical Institute. He is a keen pilot and diver.

MORE TO EXPLORE

The Fabric of Reality: The Science of Parallel Universes—and Its Implications. David Deutsch. Penguin, 1998.

The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation. Dirk Bouwmeester, Artur Ekert and Anton Zeilinger. Springer, 2000.

Quanta, Ciphers and Computers. Artur Ekert in *The New Physics*. Edited by Gordon Fraser. Cambridge University Press, 2006.

The Beginning of Infinity: Explanations That Transform the World. David Deutsch. Penguin, 2011.

The Emergent Multiverse: Quantum Theory according to the Everett Interpretation. David Wallace. Oxford University Press, 2012.

© 2013 Scientific American

Q&A SCIENCE TALK

EXTREME PHYSICS

PROFILE

BAD BOY OF PHYSICS

Leonard Susskind rebelled as a teen and never stopped. Today he insists that reality may forever be beyond reach of our understanding

Interview by Peter Byrne

IN BRIEF

WHO
LEONARD SUSSKIND

VOCATION | AVOCATION
Theoretical physicist, known especially for pioneering string theory, black hole physics and the multiverse

WHERE
Stanford University

RESEARCH FOCUS
What is the deep nature of physical reality?

BIG PICTURE
We may never be able to grasp that reality. The universe and its ingredients may be impossible to describe unambiguously.

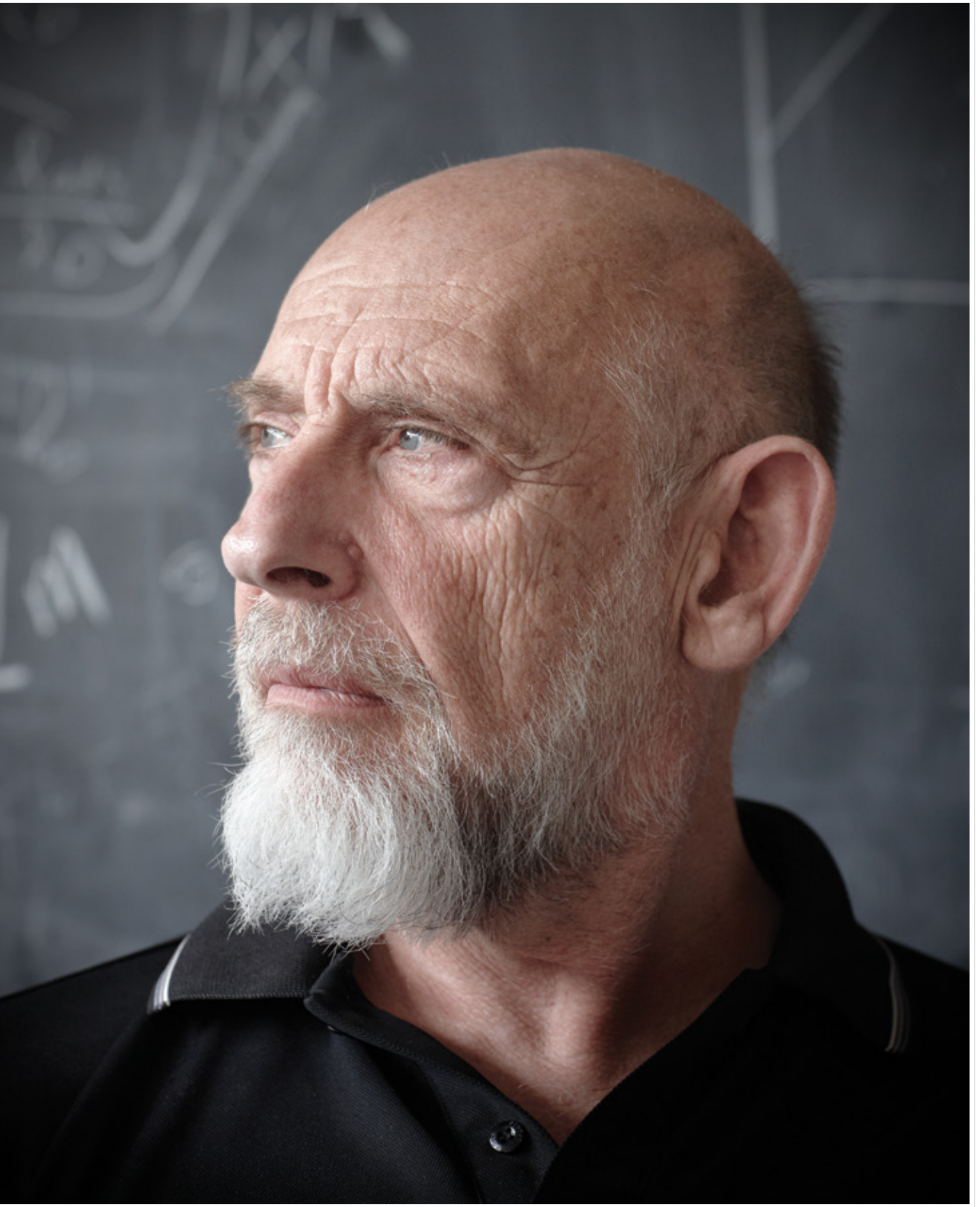
STANFORD UNIVERSITY PHYSICIST LEONARD SUSSKIND revels in discovering ideas that transform the status quo in physics. Forty years ago he co-founded string theory, which was initially derided but eventually became the leading candidate for a unified theory of nature. For years he disputed Stephen Hawking's conjecture that black holes do not merely swallow objects but grind them up beyond recovery, in violation of quantum mechanics. Hawking eventually conceded. And Susskind helped to develop the modern conception of parallel universes, based on what he dubbed the "landscape" of string theory. It spoiled physicists' dream to explain the universe as the unique outcome of basic principles.

Physicists seeking to understand the deepest levels of reality now work within a framework largely of Susskind's making. But a funny thing has happened along the way. He now wonders whether physicists *can* understand reality.

Susskind worries that reality might be beyond our limited capacity to visualize it. He is not the first to express such a concern. In the 1920s and 1930s the founders of quantum mechanics split into realist and antirealist camps. Albert Einstein and other realists held that the whole point of physics is to come up with some mental picture, however imperfect, of what objective reality is. Antirealists such as Niels Bohr said those mental images are fraught with peril; scientists should confine themselves to making and testing empirical predictions. Susskind thinks the contradictions and paradoxes of modern physics vindicate Bohr's wariness.

One thing that led Susskind to this conclusion is his principle of black hole com-

© 2013 Scientific American



Photograph by Timothy Archibald

© 2013 Scientific American

plementarity, which holds that there is an inherent ambiguity in the fate of objects that fall into a black hole. From the point of view of the falling object itself, it passes without incident through the hole's perimeter, or horizon, and is destroyed when it reaches the hole's center, or singularity. But from the vantage point of an external observer, the falling object is incinerated at the horizon. So what really happens? The question, according to the principle of black hole complementarity, is meaningless: both interpretations are valid.

A related idea favoring antirealism is the holographic principle that Susskind and Dutch Nobel laureate Gerard 't Hooft of Utrecht University formulated in the mid-1990s. It holds that what happens in any volume of spacetime can be explained by what happens on its boundary. Although we usually think of objects as zipping around three-dimensional space, we can equally well think of them as flattened blobs sliding across a two-dimensional surface. So which is the true reality: the boundary or the interior? The theory does not say. Reality, in this holographic conjecture, is perspectival.

Hoping to better understand how the tension between hard evidence and unproved conjecture works at the frontier of physics, we asked Susskind to explain how his ideas have evolved.

Scientific American: How did the son of a Bronx plumber end up questioning the nature of reality?

LEONARD SUSSKIND: I was a bad high school student. I was very good in mathematics, but I was a bad boy, and I got in trouble a lot. The effect of that is I wasn't allowed to take regular physics. I was told I had to take automotive physics. But then in college, which was an engineering school, I took my first physics course. I was just so much better than anybody else, including the professor. And fortunately, it was not a source of contention between us that I could do the things he couldn't. But then I was actually told by one of the engineering professors that he didn't think I was cut out to be an engineer, which was correct. I asked him, "What should I do?" He said, "Well, you're exceptionally smart. You should become a scientist."

Did you take any philosophy courses?

Yeah, I did in college. I was quite fascinated by some of the concepts. My interest in it lapsed when I really got hooked by physics.

Are there any philosophers of science whom you like?

I'm one of the few physicists I know who likes Thomas Kuhn. He was partly a historian of science, partly a sociologist. He got the basic idea right of what happens when the scientific paradigm shifts. A radical change of perspective suddenly occurs. Wholly new ideas, concepts, abstractions and pictures become relevant. Relativity was a big paradigm shift. Quantum mechanics was a big paradigm shift. So we keep on inventing new realisms. They never completely replace the old ideas, but they do largely replace them with concepts that work better, that describe nature better, that are often very unfamiliar, that make people question what is meant by "reality." Then the next thing comes along and turns that on its head. And we are always surprised that the old ways of thinking, the wiring that we have or the mathematical wiring that we may have created, simply fail us.

In the midst of all this remodeling, is there room for such a thing as an objective reality?

Every physicist must have some sense that there are objective things in the world and that it's our job to go and find out what those objective things are. I don't think you could do that without having a sense that there is an objective reality. The evidence for objectivity is that experiments are reproducible. If you kick a rock once, you'll hurt your toe. If you kick a rock twice, you'll hurt your toe twice. Do the same experiment over and over with a rock, and you'll reproduce the same effect.

That said, physicists almost never talk about reality. The problem is that what people tend to mean by "reality" has more to do with biology and evolution and with our hardwiring and our neural architecture than it has to do with physics itself. We're prisoners of our own neural architecture. We can visualize some things. We can't visualize other things.

Einstein's abstract, four-dimensional geometry was hard to concretely visualize. It became visualizable through mathematical relations. When relativity suddenly appeared, it must have seemed to many people: What happened to "real" time? What happened to "real" space? It just got mixed up into this funny thing, but there were rules. The point was there were clear and precise mathematical rules that had been abstracted out of it, and these survived, and the old notions of reality went away.

So I say, let's get rid of the word "reality." Let's have our whole discussion without the word "reality." It gets in the way. It conjures up things that are rarely helpful. The word "reproducible" is a more useful word than "real."

What about quantum mechanics? According to that theory, kicking the same rock the same way can actually give different results.

That's the big one, isn't it? There are two things that were discovered in quantum mechanics that upset our classical sense of reality. One was entanglement. What entanglement said was something very bizarre: that you can know everything there is to know about a composite system and yet not know everything about the individual constituents. It is a good example of how we're simply not biologically equipped for abstraction and how our sense of reality gets upset.

The other thing that really hit hard on the idea of classical reality was the Heisenberg uncertainty principle. If you try to describe an object as having both a position and a momentum, you'll run into trouble. You should think of it as having a position or a momentum. Don't try to do both.

This is what physicists mean by "complementary"?

Exactly. It turns out that the mathematics of the event horizon of a black hole is very similar to the uncertainty principle. Again, it's a question of "or" versus "and." At a completely classical level, something

© 2013 Scientific American

BLACK HOLES reveal the limits of our capacity to understand the universe.

DON DIXON

falls into a black hole, something doesn't fall into a black hole, whatever. There are things outside the black hole, and there are things inside the black hole. What we learned is that's the wrong way to think. Don't try to think of things happening outside the horizon and things happening inside the horizon. They're redundant descriptions of the same thing. You describe it one way, or you describe it the other way. This means we have to give up the old idea that a bit of information is in a definite place.

If I get you correctly, the holographic principle extends the complementary model of a black hole to the universe.

Yes. Suppose we want to describe some system with enormous precision. To probe with great precision, you need high energy. What's eventually going to happen as you try to get more and more precise is you're going to start creating black holes. The information in a black hole is all on the surface of the black hole. So the more and more refined description you make of a system, you will wind up placing the information at a boundary.

There are two descriptions of reality: either reality is the bulk of spacetime surrounded by the boundary, or reality is the area of the boundary. So which description is real? There is no way to answer that. We can either think of an object as an object in the bulk space or think of it as a complicated, scrambled collection of information on the boundary that surrounds it. Not both. One or the other. It's an incredibly scrambled mapping of one thing to the other thing.

The original goal of string theory was to provide a unique explanation of reality. Now it gives us multiple universes. What happened?

A large fraction of the physics community has abandoned trying to explain our world as unique, as mathematically the only possible world. Right now the multiverse is the only game in town. Not everybody is working on it, but there is no coherent, sharp argument against it.

In 1974 I had an interesting experience about how scientific consensus forms. People were working on the as yet untested theory of hadrons [subatomic particles such as protons and neutrons], which is called quantum chromodynamics, or QCD. At a physics conference I asked, "You people, I want to know your belief about the probability that QCD is the right theory of hadrons." I took a poll. Nobody gave it more than 5 percent. Then I asked, "What are you working on?" QCD, QCD, QCD. They were all working on QCD. The consensus was formed, but for some odd reason, people wanted to show their skeptical side. They wanted to be hard-nosed. There's an element of the same thing around the multiverse idea. A lot of physicists don't want to simply fess up and say, "Look, we don't know any other alternative."

The universe is very, very big. Empirically we know it's at least 1,000 times bigger in volume than the portion that we can ever see. The success of the concept of cosmic inflation opens the possibility that the universe is varied on big enough scales. String theory provides Tinkertoy elements that can be put together in an enormous number of ways. So there's no point in looking for explanations of why our piece of the world is exactly the way it is because there are other pieces of the world that are not exactly the same as ours. There can't be a universal explanation of everything that it is, any more than there can be a theorem that says the average temperature of a planet is 60 degrees Fahrenheit. Anyone who tried to make a calculation to prove that planets have a temperature of 60 degrees would be foolish because there are lots of planets out there that don't have that temperature.

But nobody knows the underlying rules for multiverses. It's a picture. Nobody knows how to use this predictively. This process of eternal inflation just produces bubble after bubble after bubble and produces any number of them of every kind. So that means that the probability for one versus the other is infinity over infinity. We would like to have a probability distribution that would say one is more probable than the other and then make a prediction. So we've gone from what looks like a very compelling picture on the one hand to absurdly trying to measure an infinity of probabilities. If it's going to go down, it's going to go down because of that.

Is it possible to do theoretical physics and not have philosophical thoughts?

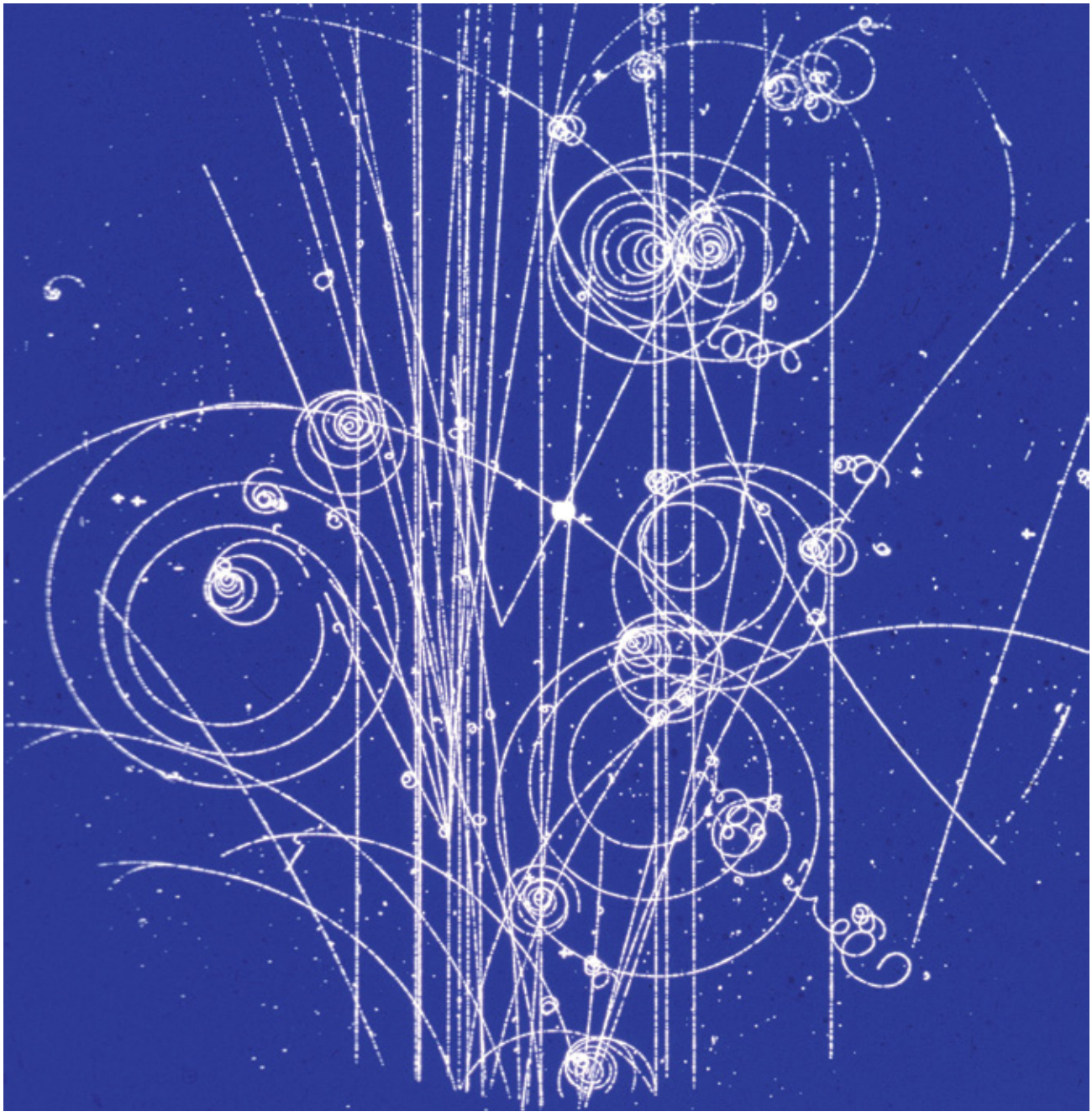
Most great physicists have had a fairly strong philosophical side. My friend Dick Feynman hated philosophy and hated philosophers, but I knew him well, and there was a deep philosophical side to him. The problems that you choose to think about are conditioned by your philosophical predispositions. But I also have a strong sense that surprises happen and put your philosophical prejudices on their head. People have the idea that there are cut-and-dried rules of science: you do experiments, you get results, you interpret them; in the end, you have something. But the actual process of science is as human and as chaotic and as contentious as anything else. SA

Peter Byrne is author of "The Many Worlds of Hugh Everett" in the December 2007 issue of *Scientific American*, which developed into the book *The Many Worlds of Hugh Everett III: Multiple Universes, Mutual Assured Destruction and the Meltdown of a Nuclear Family* (Oxford University Press, 2010).

MORE TO EXPLORE

Farewell to Reason. Paul Feyerabend. Verso, 1988.
The Cosmic Landscape: String Theory and the Illusion of Intelligent Design. Leonard Susskind. Back Bay Books, 2006.
The Black Hole War: My Battle with Stephen Hawking to Make the World Safe for Quantum Mechanics. Leonard Susskind. Back Bay Books, 2009.

© 2013 Scientific American



Signature Analysis

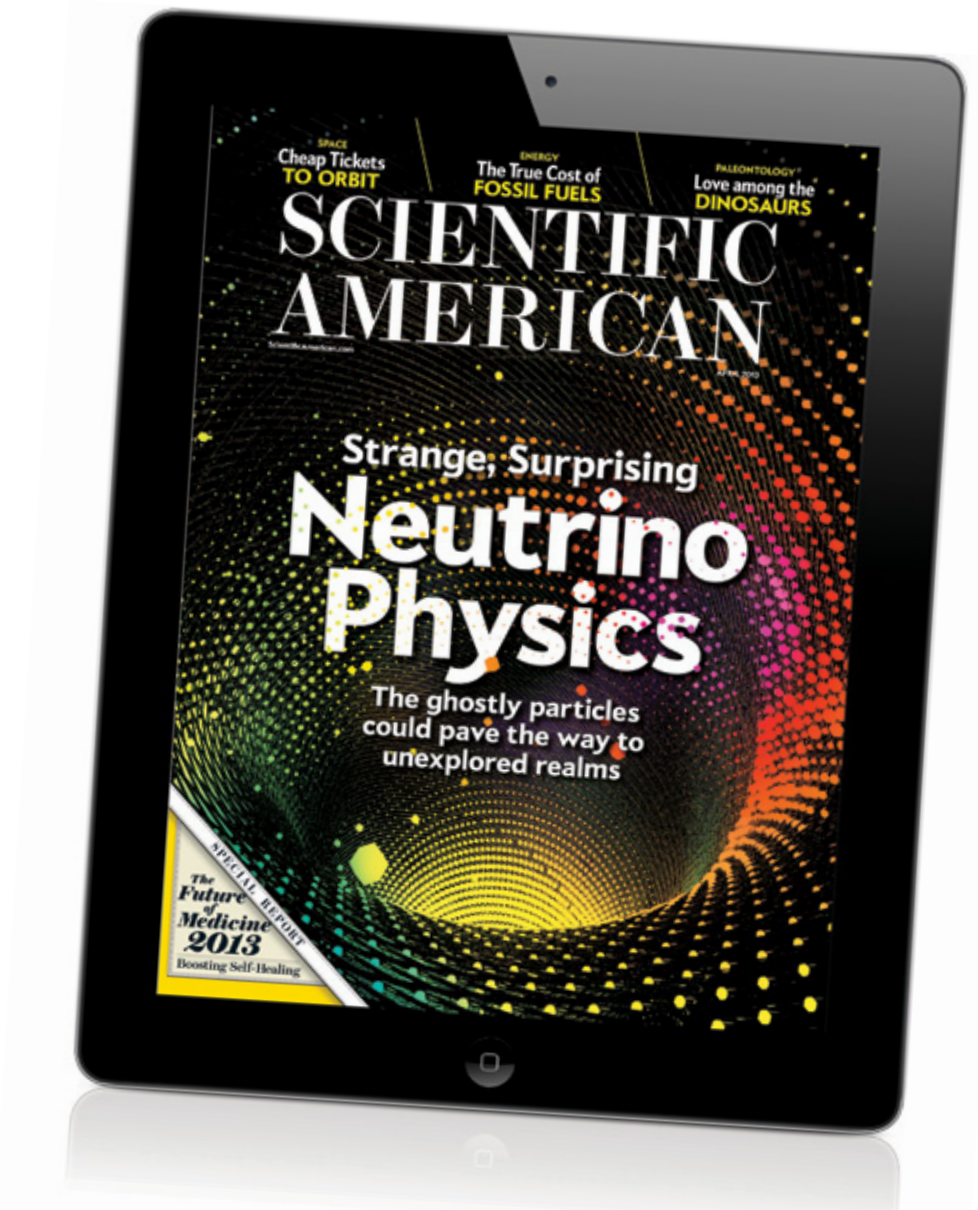
Bubble chamber photographs reveal the telltale tracks of exotic subatomic particles

Invented in 1952 by the late Donald Glaser, then at the University of Michigan, a device called a bubble chamber allows scientists to trace the paths of high-energy, charged particles rocketing through a superheated liquid. Ionizing particles leave trails made of expanding bubbles, and the shape of these trails provides clues about the identity of the particles and how they decay. For example, lightweight electrons and positrons lose energy quickly, creating spirals. Particles that elude collision leave straight, parallel tracks. This classic 1960 photograph shows tracks in the first liquid-hydrogen bubble chamber at CERN, the European laboratory for particle physics near Geneva. A negatively charged pion interacts with a proton in the liquid hydrogen and generates sprays of new particles, including a neutral lambda particle that decays into a pair of lighter, charged particles—creating the V shape seen in the center of the image. Glaser won the 1960 Nobel Prize in Physics for his invention of the bubble chamber, which led to the discovery of new particles. These days physicists view particle tracks using monster electronic detectors that can observe hundreds of millions of events per second. *—The Editors*

COURTESY OF CERN

© 2013 Scientific American

Haven't seen our tablet edition? How strange...



Get closer to the leading edge of science and technology: experience the SCIENTIFIC AMERICAN Tablet Edition - available for iPad®.

Available on the App Store

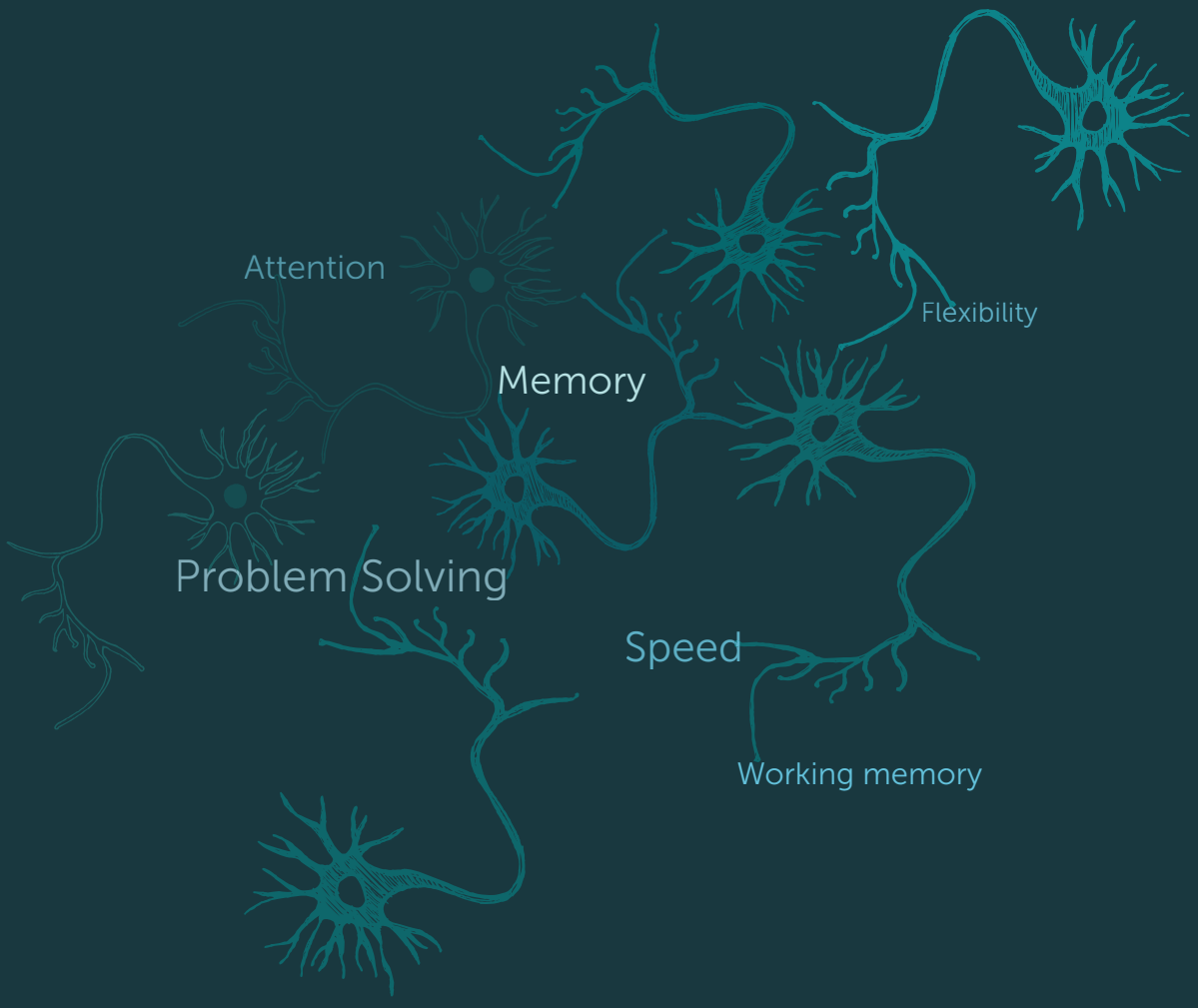
TO RENEW

SCIENTIFICAMERICAN.COM/UPGRADE

TO SUBSCRIBE

SCIENTIFICAMERICAN.COM/TABLET

Apple and IPad are trademarks of Apple Inc., registered in the U.S. and other countries. App Store is a service mark of Apple Inc.



In 2013, unlock the secrets of your brain.

With **35 million users** worldwide and the **largest database** on human cognition ever assembled, Lumosity's brain training program helps leading **research collaborators** understand the human brain.

Now **you** can advance neuroscience research and unlock your own amazing potential in the process: sign up for Lumosity training today.

Every user's progress helps neuroscientists understand the brain a little better. That's how Lumosity makes the world a brighter place.

Unlock your potential at **lumosity.com**

lumosity